



COLLEGE PARK CAMPUS

**DISPERSION ANALYSIS AND ERROR ESTIMATION  
OF GALERKIN FINITE ELEMENT METHODS  
FOR THE NUMERICAL COMPUTATION OF WAVES**

by

**Frank Ihlenburg**

and

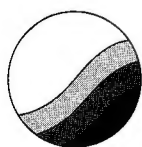
**Ivo M. Babuška**



**Technical Note BN-1174**

19950131 078

**July 1994**



INSTITUTE FOR PHYSICAL SCIENCE  
AND TECHNOLOGY

**DISTRIBUTION STATEMENT A**

Approved for public release;  
Distribution Unlimited

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Note BN-1174	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Dispersion Analysis and Error Estimation of Galerkin Finite Element Methods for the Numerical Computation of Waves		5. TYPE OF REPORT & PERIOD COVERED Final Life of Contract
7. AUTHOR(s) Frank Ihlenburg <sup>1</sup> - Ivo M. Babuska <sup>2</sup>		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Institute for Physical Science and Technology University of Maryland College Park, MD 20742-2431		8. CONTRACT OR GRANT NUMBER(s) 2 N00014-93-I-0131 (ONR) & #517 402 524 3/DAAD
11. CONTROLLING OFFICE NAME AND ADDRESS Department of the Navy Office of Naval Research Arlington, VA 22217		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE July 1994
		13. NUMBER OF PAGES 48
		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)  Approved for public release: distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  The authors' recent result on the phase difference for one-dimensional problems in numerically evaluated and discussed in the context of other work directed to this topic. It is shown that previous error estimates in integral norm are of nondispersive character but hold for medium or high wavenumber on extremely refined mesh only. On the other hand, recently proven error estimates on normalized mesh contain a pollution term. With certain assumptions on the exact solution, this term is of the order of the phase difference. Thus a link is established between the results of dispersion analysis and the results of numerical analysis.		

# Dispersion Analysis and Error Estimation of Galerkin Finite Element Methods for the Numerical Computation of Waves

Frank Ihlenburg

Ivo Babuška

July 26, 1994

*Institute for Physical Science and Technology,  
University of Maryland at College Park, College Park, MD 20742*

## Abstract

When applying numerical methods for the computation of stationary waves from the Helmholtz equation, one obtains "numerical waves" that are dispersive also in nondispersive media. The numerical wave displays a phase velocity that depends on the parameter  $k$  of the Helmholtz equation. In dispersion analysis, the phase difference between the exact and the numerical solutions is investigated. In this paper, the authors' recent result on the phase difference for one-dimensional problems is numerically evaluated and discussed in the context of other work directed to this topic. It is then shown that previous error estimates in integral norm are of nondispersive character but hold for medium or high wavenumber on extremely refined mesh only. On the other hand, recently proven error estimates on normalized mesh contain a pollution term. With certain assumptions on the exact solution, this term is of the order of the phase difference. Thus a link is established between the results of dispersion analysis and the results of numerical analysis. Throughout the paper, the presentation and discussion of theoretical results is accompanied by numerical evaluation of several model problems. Special attention is given to the performance of the Galerkin method with higher order of polynomial approximation  $p$  ( $h-p$ -version).

<input checked="checked" type="checkbox"/>
<input type="checkbox"/>
<input type="checkbox"/>

By _____	
Distribution / _____	
Availability Codes	
Dist	Avail and / or Special
A-1	

DISTRIBUTION STATEMENT A
Approved for public release; Distribution Unlimited

# 1 Introduction

1. The analytical investigation of waves in nondispersive media is based on the reduced wave or Helmholtz equation

$$\Delta u + \frac{\omega^2}{c^2} u = 0 \quad (1)$$

where

- $\omega$  is the *frequency* of a particular sinusoidal (in time) wave and
- $c$  is the *speed of sound*, depending upon material properties of the medium only (not on the frequency).

The ratio  $k := \omega/c$  is called the (scalar) *wavenumber*.

It is well known that numerical solutions to the Helmholtz equation, as finite element or finite difference solutions, do not preserve the nondispersive character of the mathematical model and its exact solutions. More specifically, when applying these methods, the solutions have the form of dispersive “numerical waves” with a “discrete wavenumber”  $k^h$  that depends on the frequency or, equivalently, on the exact wavenumber  $k$ . Thus, the numerical waves propagate with a *phase velocity*  $c^h := \omega/k^h$  that, in turn, is different from the speed of sound  $c$ .

2. In dispersion analysis, the dispersive effects in numerical solutions to the wave equation are investigated; some recent publications in this field are [1, 7, 11, 19, 20], see also [21]. The conclusions drawn in these references are based on

- analytical evaluation of dispersive numerical parameters (e.g., the discrete wavenumber) and
- discussion of the results of numerical experiments on model problems in one [11, 19, 21], two [7, 20, 21] or three dimensions [1].

In applied acoustic computations, the stepsize is usually “normalized” with respect to the exact wavelength  $k$ ; thus, a “rule of the thumb” [11] in these computations is to resolve the wavelength by 10 elements (when applying the Galerkin method with standard linear elements). This amounts to the formula  $hk = \pi/5$ . Hence the stepwidth computed by the rule, but not the rule itself, depends on the frequency  $k$ . In other words, the rule is of nondispersive character whereas the numerical solution is dispersive. It is therefore of practical interest to explore the implications of numerical dispersion on the accuracy of the computed solution [1, 19].

The results of dispersion analysis have also led to the Galerkin least-squares methods that reduce or even eliminate the dispersive numerical effects by modification of the variational model – see Harari and Hughes[11], Thomson and Pinsky[19]. This topic has recently been addressed by one of the present authors in [6] .

3. The effect of dispersion in the numerical wavenumber  $k^h$  is measured either *directly*, by comparing (graphically and analytically by means of Taylor expansions)  $k^h$  to  $k$ , or *indirectly*, by comparing exact stationary waves to their numerical counterparts. In the latter case, the dispersive effect shows in the form of a *phase lead* of the numerical waves, causing significant numerical error. Both the direct and indirect dispersive effects *grow with wavenumber  $k$* , also if the mesh is normalized. It has been established that in one-dimensional computations the phase error for the standard Galerkin FEM with linear elements is of order  $\mathcal{O}(k^3 h^2)$  [11].

In [7], Bayliss et al. compute the error of the finite element solution for a two-dimensional model problem (Helmholtz equation on a square with Dirichlet and Neumann boundary conditions). It is shown in numerical experiments that on normalized mesh (with  $kh = 0.204$ ) the error measured in  $L^2$ -norm is  $\mathcal{O}(k^3 h^2)$ . In [19], Thomson and Pinsky study dispersive effects for the Galerkin FEM with

- different local approximation basis (Legendre, spectral and Fourier elements) and
- different order of local approximation ( $p = 1 \dots 5$ ).

From a Taylor analysis similar to the approach applied by Harari and Hughes [11], carried out for  $p = 1, 2, 3$ , the general formula [19, p. 267]

$$\frac{k^h}{k} - 1 = \mathcal{O}(kh)^{2p} \quad (2)$$

is conjectured. Furthermore, it is found that the choice of the approximation basis has a negligible effect on the dispersive characteristics of the numerical solution (except for slightly worse performance of spectral elements), whereas "the higher order  $p$ -elements exhibit increased accuracy compared to low-order finite elements for the same number of degrees of freedom" [19, p. 266].

4. Numerical analysis of the reduced wave equation has been, until recently, based on a well known proposition on indefinite variational forms, shown by Schatz in 1974 [16]. This proposition states that, provided the mesh is sufficiently fine, the error of the finite element solution is essentially (i.e. up to a constant) equal to the error of the best approximation in the subspace that is determined by the choice of the FEM (i.e. of the stepsize  $h$  and the degree of approximation  $p$ ). This best approximation is *the* wave closest to the exact solution in the integral norm under consideration. It can be numerically computed in the space of piecewise linear, quadratic, etc. approximation *if the exact solution is known*. In particular, for one-dimensional problems the error of best approximation in the  $H^1$ -seminorm is just the interpolant to the exact solution (cf. [11]). Consequently, no phase error is present in this numerical wave: the best approximation, and hence the finite element solution on sufficiently small mesh, are nondispersive.

On first glance, this result, which also holds for Helmholtz-type variational forms, seems to contradict the findings of dispersion analysis, in particular, the numerical results

given by Bayliss et al. [7]. However, the reasons for the difference in both statements can be found in more detailed formulations of Schatz' proposition as applied to the Helmholtz equation – see Aziz et al. [2], Douglas Jr. et al. [10]. As it is shown in these papers, Schatz' proposition holds under the assumption that  $k^2h$  is small. In the terminology of dispersion analysis one can say that the dispersive effect in the error is negligible on "square-normalized" mesh, i.e. meshes with  $k^2h \equiv \text{const}$ .

For small  $k$ , where normalized and square-normalized meshes are of comparable density, no significant dispersion effect will be observed also on normalized mesh – which is consistent with the results of dispersion analysis – see, e.g., [19, p. 263], Fig. 3a.

For medium or large  $k$ , the assumptions of Schatz' theorem are outside the range of practical application, hence the predicted effect has not been observed in the dispersion analysis on normalized mesh.

5. This research had been initiated by observations made while solving different benchmark problems in fluid-solid interaction. Comparing computational results to either analytical solutions or experimental measurements it was found that the values deviated from each other above the low frequency band. Excluding computational truncation as a significant contributor to the errors, the conclusion was that the rule to normalize the stepsize and the frequency by maintaining a relation  $hk = \text{const}$  is not sufficient to control the error of the standard Galerkin method in the medium and high frequency bands.

In [14], [15], an analytical study of the Galerkin finite element method – both with piecewise linear and piecewise polynomial approximation – is presented. Unlike previous results of numerical analysis that hold on square-normalized meshes only (in the *asymptotic range* of  $h$ ), all statements in these papers are formulated for normalized meshes (*preasymptotic range*). The analysis is carried out on a one-dimensional model problem for the exterior Helmholtz equation with Dirichlet and Robin boundary conditions. While in [14] the analysis is restricted to the standard approach with piecewise linear approximation ( $h$ -version with approximation order  $p = 1$ ), the second part [15] deals with arbitrary order of approximation. As to be expected, numerical estimates are obtained that *reflect the dispersive character* of the finite element solutions. The dispersive character is reflected by a *pollution term* that is added to the approximation error in the estimates on normalized mesh.

It can be shown that the reduction of the phase error attempted by the Galerkin least squares concept is equivalent to reducing the pollution present in the finite element error when measured in integral norm [6]. The analysis in [6] also shows that in one dimension it is possible to entirely eliminate the phase error without sacrificing the optimal order of convergence.

However, in two dimensions it is not possible to *eliminate* the pollution in the finite element error by *any* modification of the Galerkin approach (cf. [6], Theorem 3.7). Still, the adverse effects of the pollution can be reduced by suitable modification of the standard method. Again, this conclusion from the numerical analysis of the problem closely corresponds to recent findings from dispersion analysis of the two-dimensional problem



[20].

6. In this paper, the main results of [14], [15] are discussed in the context of dispersion analysis, focusing on conclusions for computational application. Special attention is given to the specifics of the application of the  $h - p$ -version of the Galerkin FEM, as compared to the  $h$ -version. Thus, by reviewing, illustrating, and expanding the theoretical work, a full analytical and numerical explanation of the error behaviour throughout the range of convergence of the finite element solution is given for one-dimensional problems. The practically important question is whether this theoretical analysis is relevant to applications in two or three dimensions. To the knowledge of the authors, finite element error estimates for Helmholtz problems in two or three dimensions have not yet been proven. A first numerical assessment of this question is contained in the end of this paper. The results, as well as previous numerical tests carried out by Bayliss et al. [7], show the same error behaviour as observed in one-dimensional computations suggesting the conjecture that the one-dimensional study is well suited to provide a basic understanding of the error behaviour in two and, possibly, three dimensions. A more detailed analysis of higher-dimensional problems will be given in a forthcoming paper.

The paper is organized as follows: In section 2, the dispersive properties of numerical solutions to the reduced wave equation are analyzed and interpreted. In particular, a result from [15] on the estimation of the phase lead in Galerkin finite element solutions to the ordinary Helmholtz equation is evaluated and compared to previous results of dispersion analysis.

The focus turns to integral error estimates in section 3. Here, previously known estimates on fine mesh and the recently proven estimates of [14], [15] are discussed in the light of the dispersion analysis given in section 2. In particular, it is shown that the finite element solutions, both for the  $h$ - and the  $h - p$ -versions of the Galerkin FEM, are numerically polluted in the preasymptotic range. The connection to dispersion analysis lies in the fact that the pollution terms are of the same size as the dispersive term in the frequency of the numerical solution. Also in section 3, error estimates in  $L^2$ -norm are given. Numerical results are presented for three model problems: the exterior problem (1D) with Dirichlet condition at  $x = 0$ , a reduced fluid-structure interaction problem (1D) and a two-dimensional Helmholtz problem on the unit square with Robin conditions. The main conclusions of the investigation are collected once again in section 4.

## 2 Dispersive characteristics of numerical solutions

**Model problem:** Consider first a simple one-dimensional model problem – the ordinary Helmholtz equation with Dirichlet and Robin boundary conditions; cf. [2], [11], [14],[15]. In one dimension, the Robin boundary condition is an exactly absorbing boundary condition that is obtained from the Sommerfeld radiation condition, given at infinity, via introduction of an artificial boundary at finite range and subsequent Dirichlet-to-Neumann (DtN) mapping – cf. [11] and references therein.

Let  $\Omega = (0, 1)$  and :

$$u''(x) + k^2 u(x) = -f(x) \quad (3)$$

with boundary conditions

$$u(0) = 0 \quad (4)$$

$$u'(1) - iku(1) = 0. \quad (5)$$

Equivalently, one solves the variational problem: Find  $u \in H^1(\Omega)$ ,  $u(0) = 0$  such that

$$B(u, v) = (u', v') - k^2(u(x), v) - iku(1)\bar{v}(1) = (f(x), v), \quad (6)$$

where the brackets denote the integral product

$$(u, v) := \int_{\Omega} u(x)\bar{v}(x)dx,$$

holds for all  $v \in H^1(\Omega)$ ,  $v(0) = 0$ . Here,  $H^1$  is the Sobolev space of all functions that are square-integrable together with their first derivatives (in the sense of distributions) – see [14], [15].

Since the discussion is focused on the case of medium to large wavenumber  $k$ , it is assumed that  $k \geq 1$ .

Results of the dispersion analysis for this problem are given in [11]. A closely related example, namely, a Dirichlet fixed bar (leading to the solution of the Helmholtz equation with Dirichlet conditions at both boundaries), is considered in [19]. The ordinary Helmholtz equation with Robin/ Robin boundary conditions is analyzed in [10], where the asymptotic estimate on square-normalized mesh (see introduction) is proven.

In [2], it is shown for the model problem (6) that the Schatz proposition of quasi-optimality holds for the Helmholtz equation with the assumption  $k^2 h < c^*$ , where  $c^*$  is a sufficiently small constant ([2], Theorem 3.).

Using a more general model formulation, where via additional parametrization of both the equation and the boundary conditions the modeling of coupled media is included, the problem has been investigated by Demkowicz [9]. In this last mentioned paper, the computation and the parametrical study of analytical and discrete eigenvalues and inf-sup-constants is the prior objective of investigation.



**Discrete wavenumber:** For the numerical solution of the model problem, a discrete uniform mesh consisting of  $n + 1$  nodes,  $X_h = \{x_0 = 0, x_1 = h, \dots, x_n = 1\} \subset \Omega$ , is introduced. As usual, the parameter  $h = 1/n$  is called the *stepsize* of the mesh and the intervals  $\Delta_i = (x_{i-1}, x_i)$  are called *finite elements*. One seeks approximate solutions that are written within each finite element as polynomials of order  $p$ . The parameter  $p$  is called the *approximation degree* of the finite element solution. Thus, the finite element solution depends on three parameters: the parameter  $k$  of the equation itself and the numerical parameters  $h$  and  $p$ .

Solving the model problem by standard Galerkin finite element approach on uniform (in  $h$  and  $p$ ) mesh, one arrives at the discrete system of linear algebraic equations [14], [15]

$$[L_h] \{u_h\} = \{R_h\} \quad (7)$$

where

- the discrete operator

$$[L_h] = \begin{bmatrix} 2S_p(K) & T_p(K) & & & \\ T_p(K) & 2S_p(K) & T_p(K) & & \\ & & \ddots & & \\ & T_p(K) & 2S_p(K) & T_p(K) & \\ & & T_p(K) & S_p(K) - iK & \end{bmatrix},$$

is a complex  $n \times n$ -matrix,

- $\{u_h\}$  is the vector of the nodal values of the finite element solution  $u_{fe}$  on  $X_h$  and
- $R_h$  is the vector of the discrete right hand side.

If  $p > 1$ , the system (7) is obtained by the method of static condensation (see [15] for details), provided the *normalized frequency*  $kh = \omega h/c$  is such that this process is well-defined.

From the fundamental system of (7) one derives [11], [14]

$$\cos k^h h = -\frac{S_p(K)}{T_p(K)} \quad (8)$$

which determines the *discrete wavenumber*  $k^h$  as a function of the parameters  $k$ ,  $h$  and  $p$ . Here,  $K = kh = \omega h/c$  denotes the normalized frequency.  $S_p$  and  $T_p$  are, in general, rational polynomial functions of  $k$ . For  $p = 1$ , e.g., one derives [14]

$$S(K) = 1 - \frac{K^2}{3} \quad T(K) = -1 - \frac{K^2}{6}.$$

For higher  $p$ , the functions  $S$  and  $T$  formally depend also on the polynomial basis that is used for the  $p$ -approximation. In the numerical examples, Legendre-based,  $p$ -hierarchical elements are used – for details see Szabó and Babuška [18]. Solving eq (8) for  $k^h$ , one exhibits two dispersive phenomena characterizing the numerical wave: the cutoff frequency and the phase lead w.r. to the exact wave.

**Cutoff frequency:** For certain values of the normalized frequency  $K$ , namely, when  $|S_p(K)/T_p(K)| > 1$ , the discrete wavenumber  $k^h$  is complex with a non-vanishing imaginary part (note that the exact wavenumber  $k$  is real by assumption).

In Fig. 1, the function  $\cos(k^h h) = -S_p(K)/T_p(K)$  is plotted as a function of the normalized frequency  $K$  for piecewise linear approximation ( $p = 1$ ).

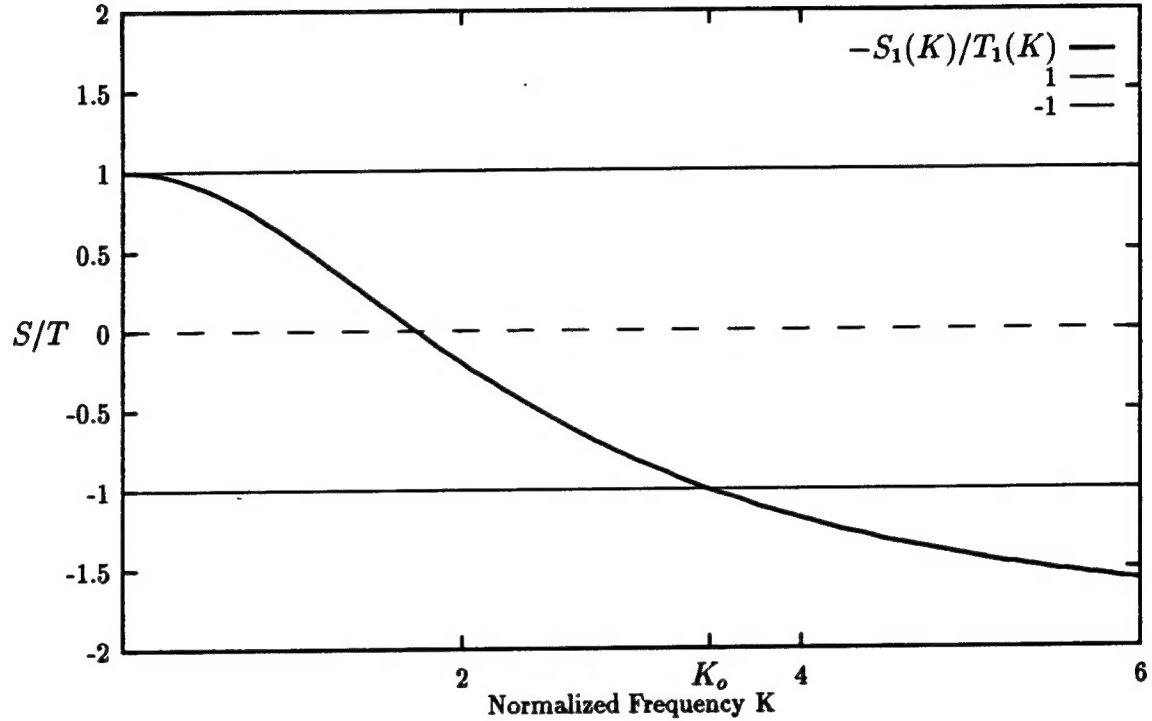


Figure 1: Cosine of the normalized discrete wavenumber vs. normalized frequency  $K$  for  $p = 1$  with cutoff frequency  $K_o$ .

At  $K_o = \sqrt{12}$  one has  $|S(K_o)/T(K_o)| = 1$ . For  $K \leq K_o$ , the discrete wavenumber  $k^h$  is real whereas it is fully complex for  $K > K_o$ . If the exact solution is a propagating wave, the numerical solution is propagating only for normalized frequencies  $K$  below the *cutoff frequency*  $K_o$ ; for frequencies beyond this value, the numerical solution assumes the form of a decaying wave [11], [19].

The magnitude of the cutoff frequency depends on the order of approximation  $p$ ; generally speaking,  $K_o$  grows with the increase of  $p$ . This effect is highlighted in Fig. 2.

Before reaching the cutoff frequency, i.e. the infinite complex interval, the discrete wavenumber is fully complex also on  $(p - 1)$  finite intervals, the stopping bands [20], as can be seen in Fig. 2 for  $p = 1 \dots 6$ .

Returning to eq (8), in Fig. 3 the cosines of the normalized discrete wavenumber  $k_h$  are graphically compared, for  $p = 1, \dots, 6$ , to the cosine of the normalized frequency. The

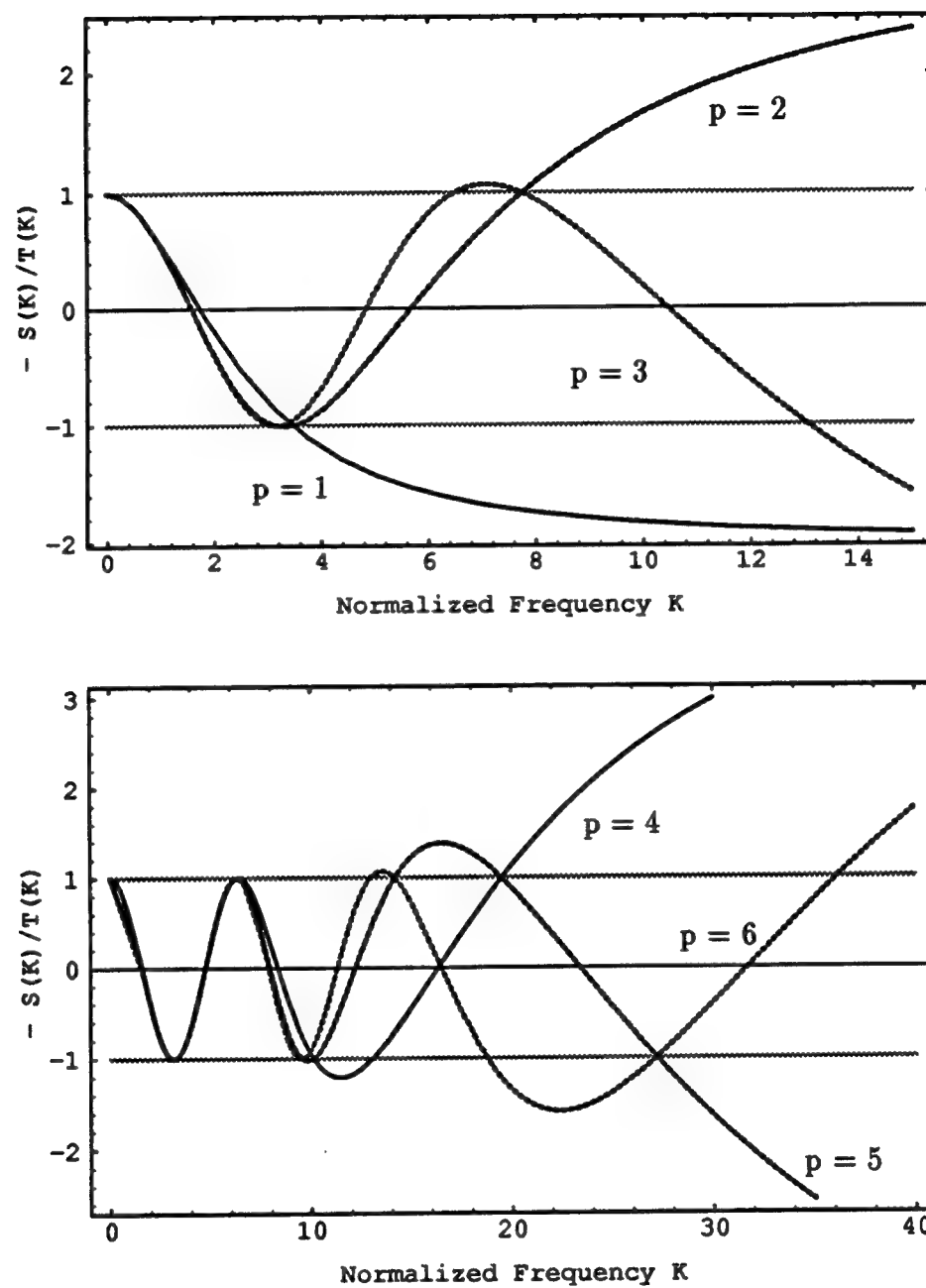


Figure 2: Cosine of the normalized discrete wavenumber vs. normalized frequency  $K$

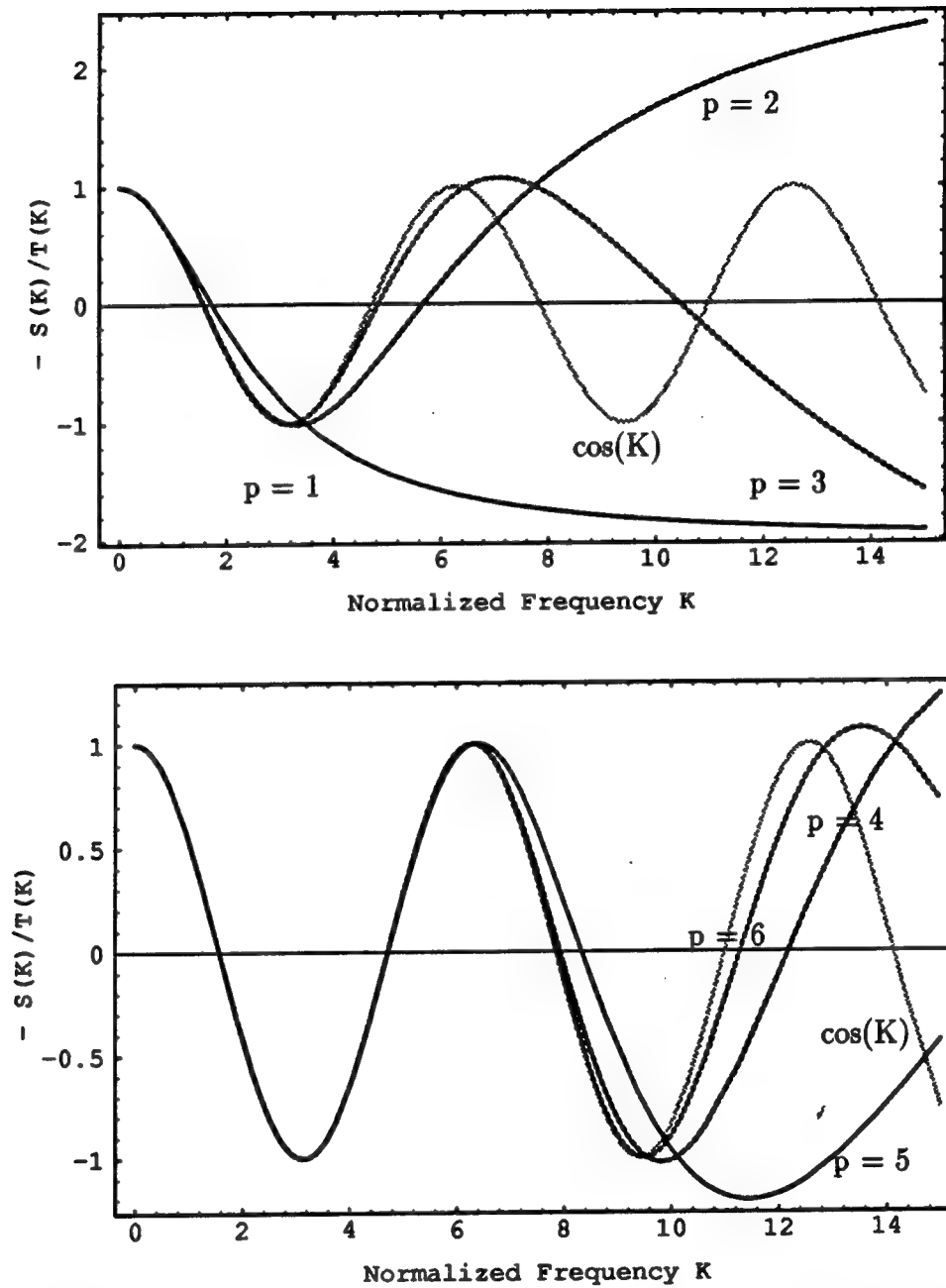


Figure 3: Cosine of the normalized discrete wavenumber vs. normalized frequency  $K$

plots show how the discrete wavenumber  $k_h$  deviates from the exact wavenumber  $k$ , and how the deviation decreases with increasing  $p$ . This phenomenon is investigated in the next paragraph.

**Phase lead of the numerical solution:** It is well known that the numerical waves display a phase lead w.r. to the exact waves – cf. Fig. 4.

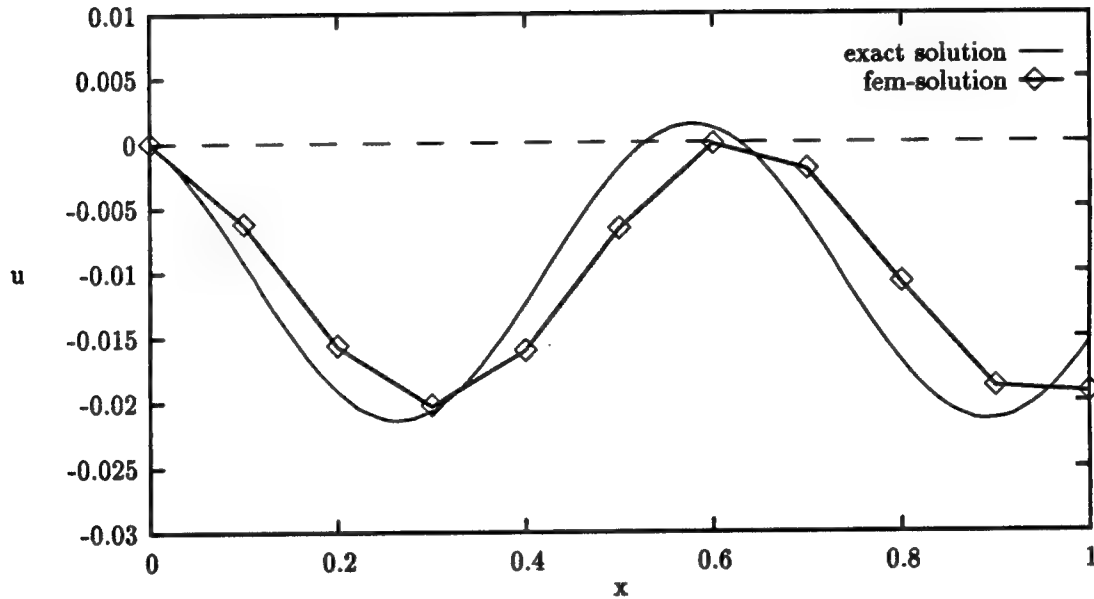


Figure 4: Finite element solution versus exact solution of eq(6) for  $k = 10, n = 10, p = 1$

On uniform mesh, the discrete wavenumber can be explicitly computed from eq (8). See in Fig. 5 results for  $k = 10, p = 1$ : the discrete wavenumber converges to the exact wavenumber as  $h$  is decreased. The theoretical rate of convergence is easily obtained from Taylor expansion in (8). For  $p = 1$ ,

$$kh - k^h h = \frac{k^3 h^3}{24} + \mathcal{O}((kh)^5). \quad (9)$$

Neglecting higher terms, one has equivalently

$$k - k^h = k \mathcal{O}((kh)^2). \quad (10)$$

This formula is generalized to the case of arbitrary  $p$  as follows: [15]

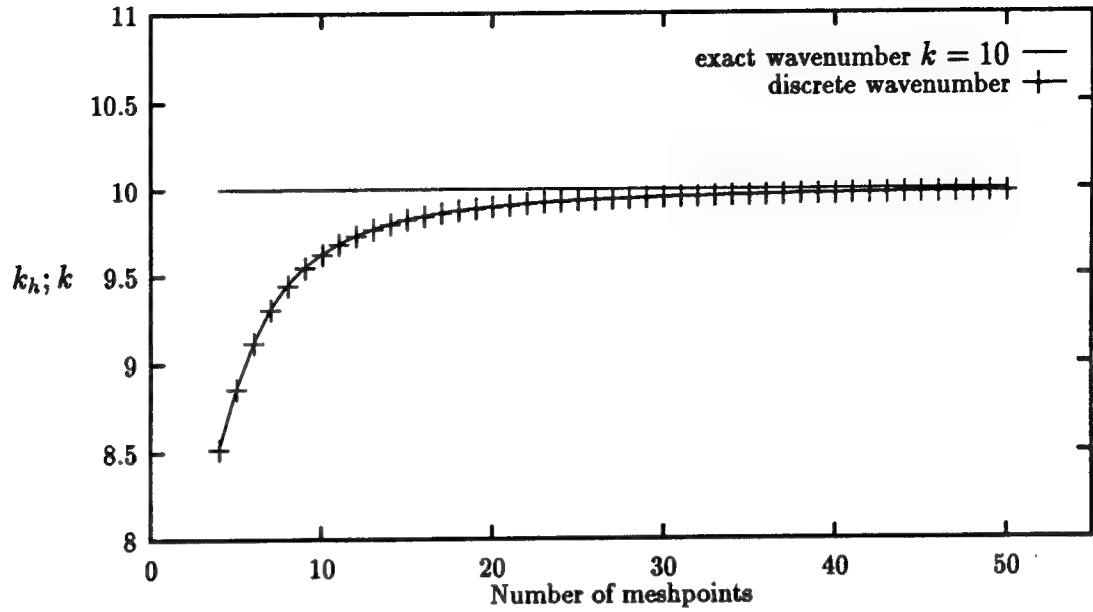


Figure 5: Discrete wavenumber  $k^h$  versus exact wavenumber  $k = \omega/c$  for  $k = 10, p = 1$  and  $n = 4 \dots 50$ .

**Theorem 1** Let  $k^h$  be the discrete wavenumber for the FEM-solution of eq (6) with wavenumber  $k$  on uniform  $h - p$  mesh.

Then, if  $hk/p < 1$ ,

$$|k^h - k| \leq k C(p) \left( \frac{hk}{p} \right)^{2p} \quad (11)$$

holds with

$$C(p) = C_1 \left( \frac{e}{4} \right)^{2p} \frac{(\pi p)^{-1/2}}{4} \quad (12)$$

where  $C_1$  is a constant not depending on  $h, k$  and  $p$ .

**Remark 1:** This statement had been induced (from computational results) as  $k^h/k - 1 = O(hk)^{2p}$  in [19]. The analytical result in Theorem 2.1 shows that the decrease of the phase difference with growing  $p$  is even more substantial – in fact, this difference decreases also "almost" by a factor  $(2p)^{2p}$ .

**Remark 2:** The estimate (11) of the phase difference between the numerical and the exact solution will be shown to be sharp by numerical evaluation of the model problem. The adverse influence of this phase lead on the accuracy of the solution is shown in Fig.1. By



estimating the size of the phase lead, one indirectly or qualitatively estimates the error of the finite element solution. Quantitative estimates of this error are furnished by error estimates in integral norms – see next section.

A graphical interpretation of estimate (11) is given in Fig. 6, where numerical results for  $\ln(k^h/k - 1) + 2p\ln(2p)$  are plotted versus  $\ln(kh)$ . One easily verifies the predicted decrease rates of the phase difference.

Checking the points of intersection of the lines one also observes that  $C$  does not significantly increase with  $p$ .

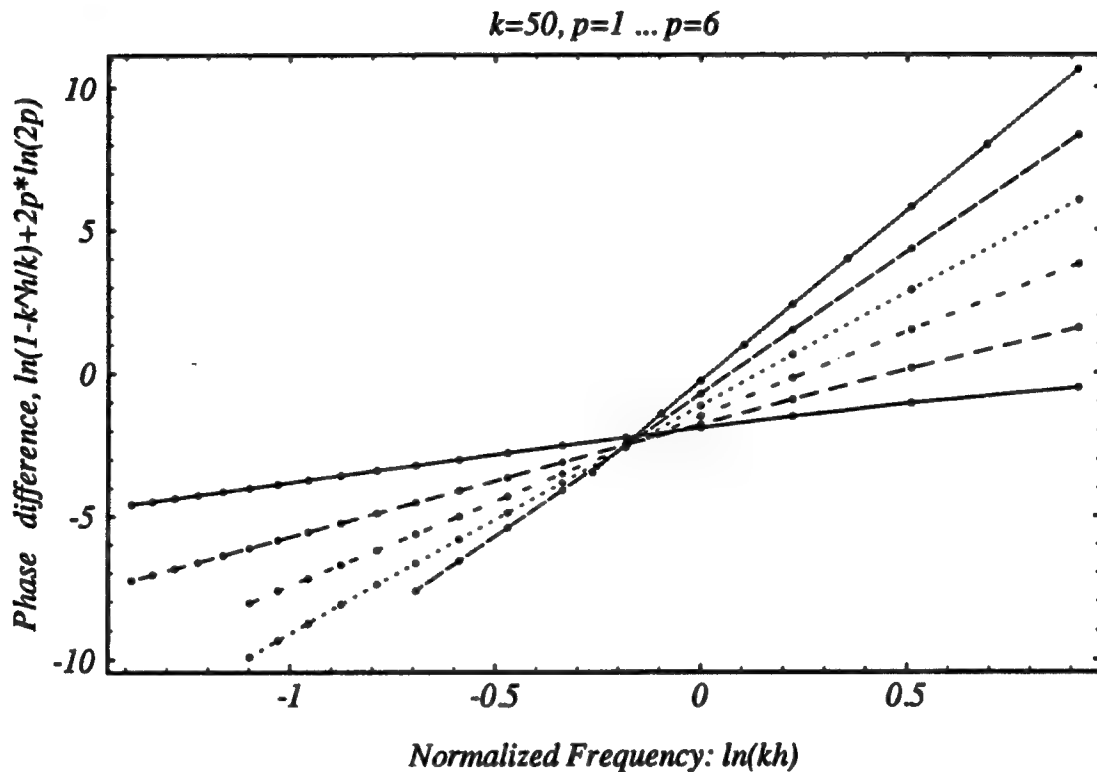


Figure 6: The phase difference as a function of normalized frequency in log-log-scale for  $k = 60$  and order of approximation  $p = 1 \dots p = 6$ .

For further interpretation of Theorem 1, the following notations are introduced.

**Definition 1** Considering finite element meshes for Helmholtz problems with parameter  $k$ , a mesh with stepsize  $h$  is called normalized if  $h$  is chosen by a constraint in the form  $hk = \alpha$  where  $\alpha$  is an a priori chosen constant independent of  $k$  and  $h$ . If a  $h-p$ -method is applied, the magnitude

$$\theta = \frac{hk}{p}$$

is called the scale of the  $h - p$ -mesh.

Obviously, for fixed  $p$ , normalization of a mesh is equivalent to fixing the scale of the mesh.

The estimate (11) states that the *normalized* phase difference on normalized mesh depends on the scale  $\theta$  and the approximation degree  $p$  only,

$$\left| \frac{kh}{p} - \frac{k^h h}{p} \right| \leq C(p) \theta^{2p+1}.$$

Indeed, when plotting, for a number of different scales, the normalized phase differences divided by  $\theta^{2p+1}$  it can be seen from Fig. 7 that the measured constant for the given example grows with  $p$ , in fact more moderately than predicted by the upper estimate (12) – find in Table 1 the computed numbers.

**Table 1:**  $h - p$ -normalized phase difference from numerical computation compared to estimate (11) with  $C_1 = 1$  for  $k = 60$ ,  $p = 1 \dots p = 6$  and  $\theta = 0.05, 0.1, 0.2$ .

$p$	$C, \text{ measured}$			$C, \text{ upper estimate}$
	$\theta = 0.1$	$\theta = 0.2$	$\theta = 0.4$	
1	.1659200470	.1637200263	.16647937140	.260551
2	.1762882103	.1718276503	.1774053010	.340336
3	.2285314750	.2199826770	.2307020	.51332
4	.3246366818	.2741436355	.3288611304	.821203
5	.486355762	.455423447	.494765034	1.35682
6	.7545962452	.696458213	1.940255003 <sup>a</sup>	2.28803

<sup>a</sup>phase difference was zero in machine precision

**Remark 3:** In the dispersion analysis of the phase difference one considers only the interior stencils of eq. (7) – see, e.g., [11, p.71]. Thus, the equation is solved discretely on the infinite interval, and the results hold independently of the boundary conditions. The same is true for Theorem 1. Note that the statement does not relate to any particular boundary conditions. The constant  $C_p$  in the estimate follows from approximation properties in the space  $H^1(\Omega)$ . Hence, though the theorem is given in [15] in the context of a particular boundary value problem, it holds equally for other one-dimensional Helmholtz problems.

**Remark 4:** The observation of the phase lead in the numerical solution has given rise to the search for modified methods which reduce this phase lead, i.e. produce a numerical wave that is closer (in phase) to the exact wave. This is achieved by the Galerkin least squares approach [11], [12], [20], [6]. In the terminology of numerical analysis, this is equivalent to the reduction of numerical pollution in the error, measured in integral norm

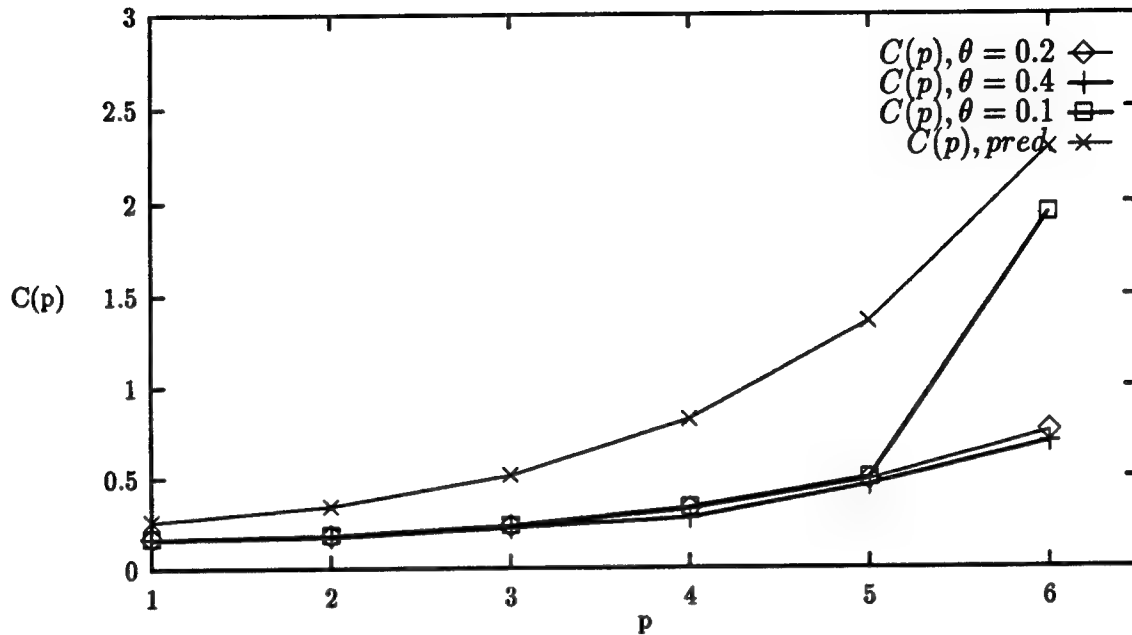


Figure 7: Upper estimate for constant  $C(p)$  versus measured values for  $p = 1 \dots 6$  and  $\theta = 0.05, 0.1, 0.2$

[6] – see next section.

It follows from the theorem that the  $h-p$ -normalized phase difference is *not dispersive* on  $h-p$ -normalized mesh with constant scale: according to eq (11) it is bounded by a magnitude that depends only on  $p$  and the scale  $\theta$ . Indeed, as to be seen from Fig. 8, the relative phase difference is indeed constant for fixed  $\theta$  and  $p$ . On the other hand, the absolute phase difference grows for fixed  $\theta$  and  $p$  linearly with  $k$ , in accord with the statement of the theorem, eq (11) – cf. Fig. 9.

The relative phase difference is obviously a local (i.e. per element) indicator of the finite element error. Since  $n$  grows with  $k$  on normalized mesh, it is intuitively clear that with higher  $k$  the global error also grows since it is obtained by adding up an increasing number of local errors. The global error should therefore expected to be dispersive.

Considering the dependence of the phase accuracy on  $p$ , estimate (11), together with eq (12) shows that for fixed  $h-p$ -scale  $\theta < 1$  and fixed  $k$ , the phase accuracy increases exponentially in both the factors  $C(p)$  and  $\theta^{2p}$  with increasing  $p$ . The theorem thus generalizes results of previous numerical studies, cf. [19], Table 1. Note that the scale is a measure for the number of degrees of freedom (DOF) of the  $h-p$ -mesh. On the

other hand, if  $k$  is increased and one wishes to control the size of the phase lag then much higher values of  $k$  can be “balanced” by the same number of DOF as  $p$  is increased. This aspect will be further commented on in the next section.

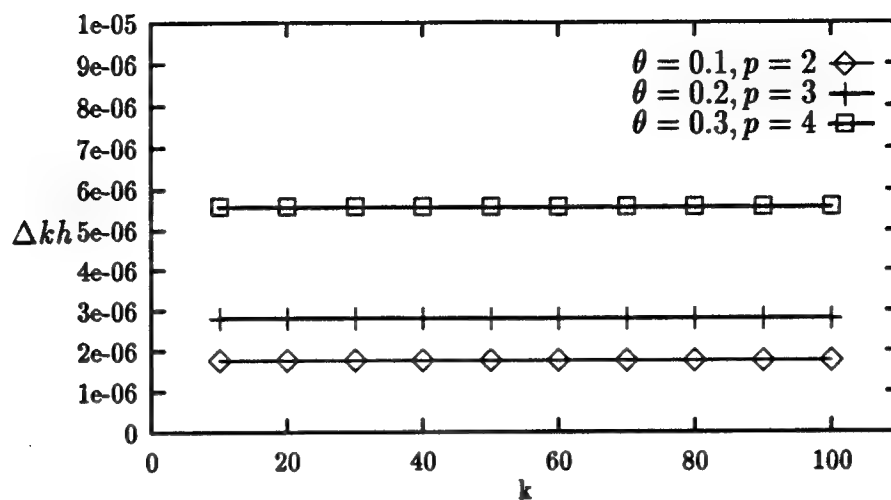


Figure 8: Relative phase difference  $(k^h - k)h/2p$  for different values of  $p$  and  $\theta$ .

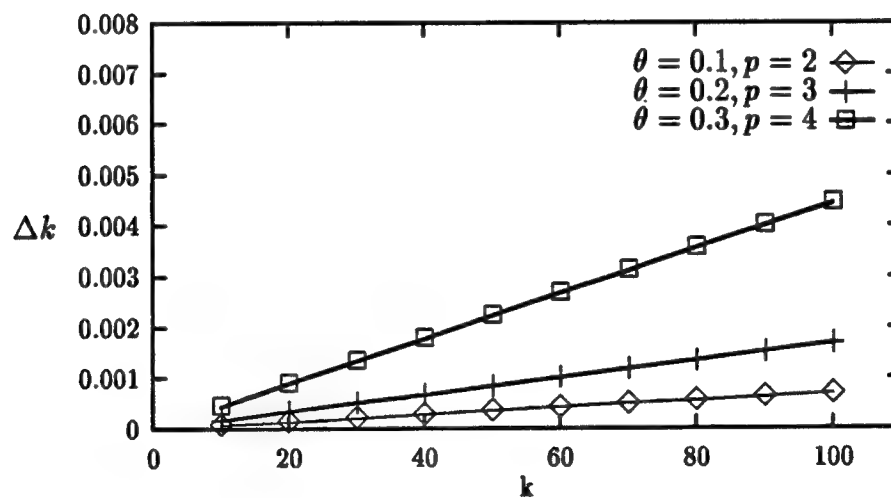


Figure 9: Absolute phase difference  $k^h - k$  for  $p, \theta$  as in Fig. 8.

### 3 Error estimates in integral norms

While dispersion analysis leads to valuable information on several physical phenomena inherent to the discrete solution and, not the least, gives qualitative insight into the sources of numerical error, it does not yield, by its nature, quantitative statements on the numerical error itself. These are furnished by error estimates that are concluded by the methods of numerical analysis. However, from the viewpoint of computational application, both methods have the same goal, namely, to provide the analyst with knowledge on the reliability of the numerical solution; it is expected that this knowledge can be used both a priori (meshdesign) and a posteriori (interpretation of the numerical solution).

In this subsection, some recently proven estimates [14], [15] are discussed in the context of dispersion analysis.

**Preliminaries:** Speaking in terms of functional analysis, the Galerkin finite element method consists in finding the solution of a variational problem – that originally is given on a normed space  $V$  – on a subspace  $V_h \subset V$ . Practically, this subspace is defined by the choice of the numerical parameters  $h$  and  $p$ :  $V_h = S_h^p(\Omega)$ .

Assuming that the variational problem has a unique solution  $u \in V$  and, if solved with restriction to  $V_h$ , also a unique solution  $u_{fe} \in V_h$ , these solutions will be called the exact and the finite element solution of the problem, resp. In general, the exact solution  $u$  does not lie in the subspace  $V_h$  and the finite element solution is an approximation of  $u$ . The error of approximation, measured in the norm of  $V$ , is

$$e_{fe} := \|u - u_{fe}\|_V.$$

Since the exact solution is generally not known in applied problems, a priori estimates of the finite element error give valuable information on the reliability of the finite element solution.

Frequently these finite element estimates are given in the form:

$$e_{fe} \leq C_{opt} \inf_{w \in V_h} \|u - w\|_V \quad (13)$$

where  $C_{opt}$  is independent of  $h$ . The infimum on the r.h.s. is the best approximation possible for functions in  $V$  by functions from  $V_h$ . Thus, via estimates of the form (13), the results of approximation theory in function spaces are directly applicable to the assessment of the finite element error.

If a finite element solution satisfies an estimate of the form (13) it is called *quasioptimal*.

**Remark 5:** For the subspaces defined by the finite element method, the infimum in (13) is reached on a uniquely defined element  $u_{ba} \in V_h$ . The error  $\|u - u_{ba}\|$  is the minimal error of approximation of  $u$  by functions from  $V_h$ . However, unlike the finite element solution  $u_{fe} \in V_h$ , the function  $u_{ba}$  is numerically computable only if the exact solution is known



analytically, i.e. not in the general case. In this investigation, where the dependence of the optimality constant on the parameters  $k$  and  $p$  is a central issue, the best approximation for the exact solution of the model problem is computed to numerically evaluate estimates given for  $C_{opt}(k, p)$ .

Specifying now the foregoing abstract discussion for the Helmholtz problems under consideration, one identifies  $V = H^1(\Omega)$  and  $V_h = S_h^p(\Omega)$ . The finite-dimensional subspaces  $V_h$  are Hilbert spaces. The norm  $\|\cdot\|_V$  is given by the Sobolev norm

$$\|v\|_1 = (\|v\|^2 + \|v'\|^2)^{1/2}$$

where

$$\|v\| = \left( \int_{\Omega} |v|^2 \right)^{1/2}$$

is the usual integral  $L^2$ -norm. Dirichlet-fixed functions can be equivalently measured in the  $H^1$ -seminorm

$$|v|_1 := \|v'\|.$$

As usual, the notation  $v^{(m)}$  is used for the  $m$ th derivative of  $v$ . If a function  $v$  is  $l$  times differentiable (in the weak sense) on  $\Omega$ , it is said to be of *regularity*  $l$ . The functions of regularity  $l \geq m > 1$  form a Hilbert subspace  $H^m(\Omega) \subset H^1(\Omega)$ . A seminorm is defined on  $H^m$  by  $|v|_m = \|v^{(m)}\|$ .

In the error estimates,  $C$  is a generic constant that does not depend on the parameters of the estimate and may have different meaning in different places of occurrence.

Detailed information on the foundations of the Finite element method can be found in [3], [8], [17].

**Error estimation on square-normalized mesh:** The notation of square-normalized mesh is given in the following definition.

**Definition 2** A finite element mesh with parameters  $h$  and  $p$  is called square-normalized if the numerical parameters are constrained w.r. to the wavenumber  $k$  by a relation

$$\frac{hk^2}{p} \leq c^*$$

where  $c^*$  is independent of  $h, k$  and  $p$ .

In [15], the following theorem is proven:

**Theorem 2** Let  $e = u - u_{f_e}$  be the error of the Galerkin finite element solution to eq (6) on uniform  $h - p$ -mesh. Assuming that the exact solution  $u$  is of regularity  $l + 1$ ,  $u \in H^{l+1}(\Omega)$ , the estimate

$$|e|_1 = \|e'\| \leq C_1(p) C_2 \left( \frac{h}{2p} \right)^m |u|_{m+1} \quad (14)$$

holds with

$$C_1(p) = \left(\frac{e}{2}\right)^p (\pi p)^{-1/4},$$

$$m = \min(l, p)$$

and

$$C_2 = \left( \frac{4 + \left(\frac{hk}{2p}\right)^2}{\frac{1}{2} - 6k^2 \left(\frac{hk}{p}\right)^2 \left(1 + \sqrt{\frac{3}{2}} \left(\frac{hk}{p}\right)^2\right)} \right)^{1/2}.$$

provided that  $h, p$  are such that the constant  $C_2$  is positive.

It is easy to see that

$$\frac{hk^2}{p} < c^* = \frac{1}{\sqrt{12}} \quad (15)$$

is necessary for  $C_2$  to be positive. On the other hand, if eq (15) is satisfied then the second term in the numerator of  $C_2$  is negligible and  $C_2$  does not depend on the parameters  $h, k, p$ . In particular, the estimate (14) is independent of the wavenumber  $k$ , i.e. of *nondispersive* character.

Thus, while the limit of resolution is  $n = k/\sqrt{12}$ , the range of nondispersive convergence theoretically begins at  $n = k^2\sqrt{12}$ .

**Remark 6:** Computational experience with the model problem shows that practically one observes optimal convergence in accord with eq (14) also on coarser mesh. However, one has, for  $p = 1$ , to maintain a scaling in the form  $k^2 h = \text{const.}$  – cf. [14]. For higher  $p$ , a weaker condition applies as will be seen below.

Theorem 2 is shown in two steps. First, one proves that

$$|u - u_{fe}|_1 \leq C_2 |u - s|_1$$

where  $s$  is the best approximation of  $u$  for given  $h$  and  $p$  (i.e. the interpolant in the present case). The statement then follows from approximation properties of the finite element spaces  $S_h^p(\Omega)$  [5] – see [15] for details.

**Example 1: Exterior problem with Dirichlet boundary condition.** For numerical evaluation, consider the model problem (6) with constant inhomogeneous data  $f \equiv 1$ . Up to scaling factors, the problem then is equivalent to the exterior Helmholtz problem with inhomogeneous Dirichlet data at  $x = 0$ . Similar problems have been evaluated by Harari/Hughes [11] and Thompson/Pinsky [19]. The analytical solution is a propagating wave

$$u(x) = \frac{1}{k^2} ((1 - \cos kx - \sin k \sin kx) + i(1 - \cos k) \sin kx). \quad (16)$$

The finite element solution is computed for  $p = 1, \dots, 6$ ; static condensation is applied if  $p \geq 2$ . Details of the solution procedure are given in [15].

The error of the finite element solution in  $H^1$ -seminorm is

$$\begin{aligned} |e_{fe}|_1^2 &= |u - u_{fe}|_1^2 \\ &= e_1^2 - \sum_{j=1}^n \frac{2}{\Delta_j} \sum_{i=3}^{p_j+1} (\bar{a}_i^{(j)} c_i^{(j)} + a_i^{(j)} \bar{c}_i^{(j)} - \bar{a}_i^{(j)} a_i^{(j)}) \end{aligned} \quad (17)$$

where

$$e_1^2 = |u|_1^2 - h \sum_{j=1}^n (D_j u D_j \bar{u}_h + D_j u_h D_j \bar{u} - D_j u_h D_j \bar{u}_h) \quad (18)$$

is the error of piecewise linear approximation (cf. [14]) Further,

$$c_i^{(j)} = \int_{\Delta_j} u' N_i'$$

and

$$D^j v = \frac{v_{j+1} - v_j}{h}$$

are the forward differences on  $X_h$ . Finally,  $a_i^{(j)}$  are the coefficients of the bubble modes in the local finite element ansatz on element  $\Delta_j$  - cf. [15].

The error of the best approximation is conveniently computed from

$$|e_{ba}| = |u|_1^2 - \sum_{j=1}^n \left( h |D^j u|^2 + \frac{2}{h} \sum_{i=3}^{p+1} |c_i^{(j)}|^2 \right) \quad (19)$$

In the case that the exact solution  $u$  is known, one can easily check numerically whether the assumption of square-normalized mesh is indeed necessary for the estimate (14) to hold. The numerical evaluation leads to an affirmative conclusion. Namely, one observes that

$$C_{opt} = \frac{|u - u_{fe}|_1}{|u - s|_1}$$

is bounded on square-normalized mesh but grows with  $k$  if a constraint  $k^\gamma h \equiv \text{const.}$  is applied with  $\gamma < 2$ .

For  $p = 1$ , these numerical results are discussed in [14].

In Figs. 10, 11, similar results are shown for  $p = 2$ . One observes in Fig. 10 that the relation  $C_{opt}$  decreases on coarse mesh for low  $k$ . This is due to the fact that, as further discussion will show, for small wavenumber the finite element solution converges with optimal rate also on normalized mesh. Equivalently, as Fig. 11 shows,  $C_{opt}$  is almost constant for low  $k$  on normalized mesh before beginning to grow oscillatory along a line  $C = \alpha n$ . Note that these examples for  $p = 2$  display the same principal effects as for  $p = 1$  - see Figs. 12, 13 from [14] for reference - but only for very high wavenumber.

In Fig. 14, the relation  $e_{fe}/e_{opt}$  is shown for  $p = 4$  with the constraint  $kh/2p = 0.5$ , i.e. more than one halfwave is resolved by one element of order 4 - cf. the discussion in [19, p. 274].

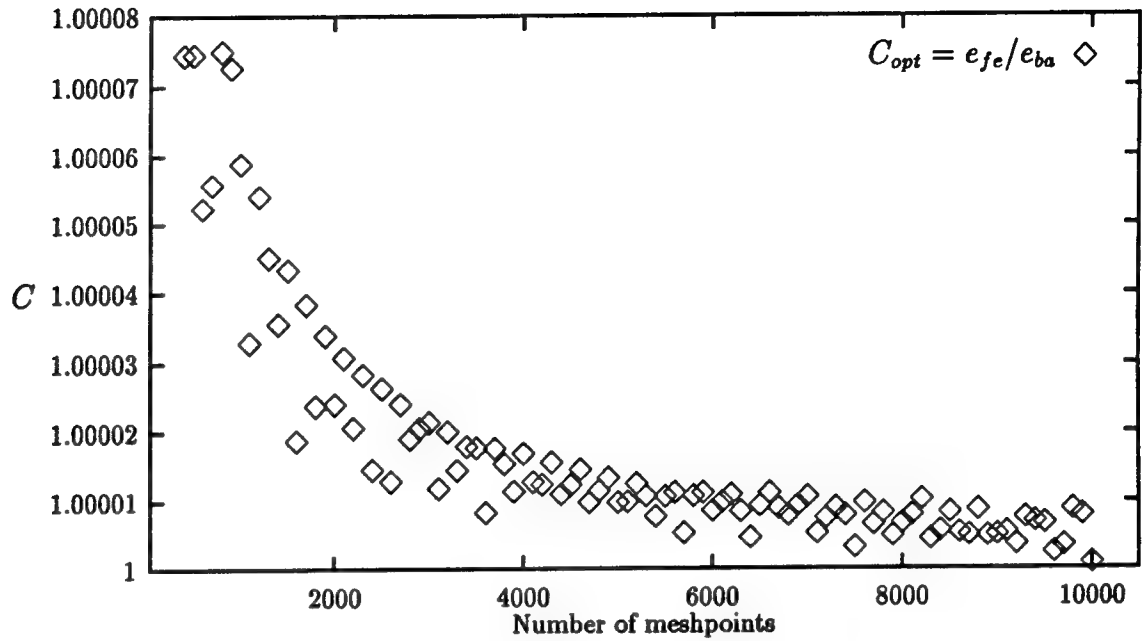


Figure 10: Relation  $e_{fe}/e_{opt}$  of the finite element error to the minimal error in the  $h-p$  approximation space in  $H^1$ -seminorm for  $p = 2$  on square normalized mesh with  $k^2 h/p = 4$

It is assumed in the following that the exact solution  $u$  is a sinusoidal stationary wave with frequency  $k = \omega/c$ . Then it is easily checked that

$$Ak^l |u|_1 \leq \|u^{(l+1)}\| \leq Bk^l |u|_1$$

where the constants  $A, B$  depend only on the amplitude of  $u$ . Solutions with this property are called *oscillatory* solutions.

Considering now the *relative* error of the FE solution, Theorem 2 yields

$$\tilde{e} := \frac{|e|_1}{|u|_1} \leq C \left( \frac{hk}{2p} \right)^p. \quad (20)$$

This means that, for fixed  $h-p$ -scale  $\theta = hk/2p \equiv \text{const.}$ , the relative error, measured in  $H^1$ -seminorm on square-normalized mesh, does not depend on  $k$ , i.e. is non-dispersive. Since it is known from numerical experience (cf., e.g. [7]), that this is not true on non-normalized mesh, the theorem cannot hold in this case. On the other hand, for large  $k$  the error almost vanishes on square-normalized mesh where the theorem does hold. Indeed, when inserting the condition

$$h = c^* \frac{p}{k^2},$$

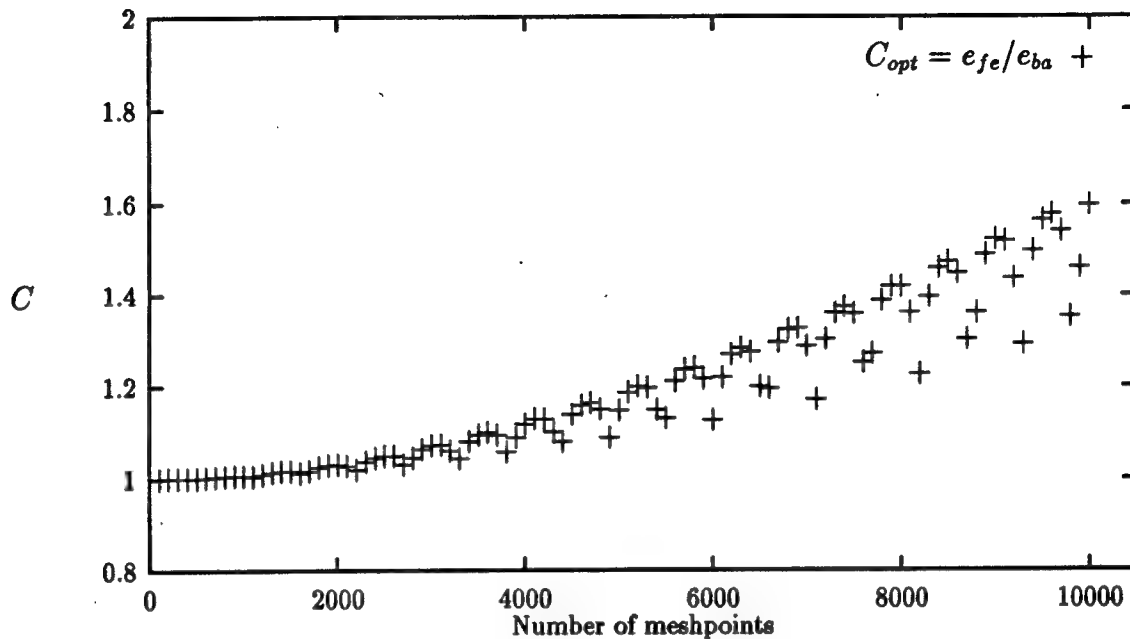


Figure 11: Relation  $e_{fe}/e_{opt}$  of the finite element error to the minimal error in the  $h - p$  approximation space in  $H^1$ -seminorm for  $p = 2$  on normalized mesh with  $hk/p = 1$ .

where  $c^*$  is some sufficiently small constant – cf. Remark 4 –, into the estimate (20), one has for any fixed  $p$

$$\tilde{e} \leq C_p \left( \frac{1}{k} \right)^p,$$

hence the error is expected to go down with rate  $k^{-1}$  on square-normalized mesh, as  $k$  is increased. This is confirmed by numerical results – see Fig. 15.

Hence, while it is practically impossible to realize square-normalization for large  $k$ , it is, on the other hand, not necessary since in computational applications the error is not expected to be close to zero but rather to stay within the limits of some tolerance.

**Error analysis on normalized mesh I:** Following the conclusion of the previous paragraph, in [14], [15] the analysis has been carried out on *normalized* mesh, i.e. under the assumption that  $hk < \alpha$ . In this case, uniqueness of the finite element solution follows from a stability estimate

$$\|u'_{fe}\| \leq C \|f\|$$

where the constant  $C$  does not depend on  $h, k$  and  $p$  or, equivalently, from a discrete inf-sup-condition ([15], Theorem 3.4). For error estimation in  $H^1$ -seminorm, the following

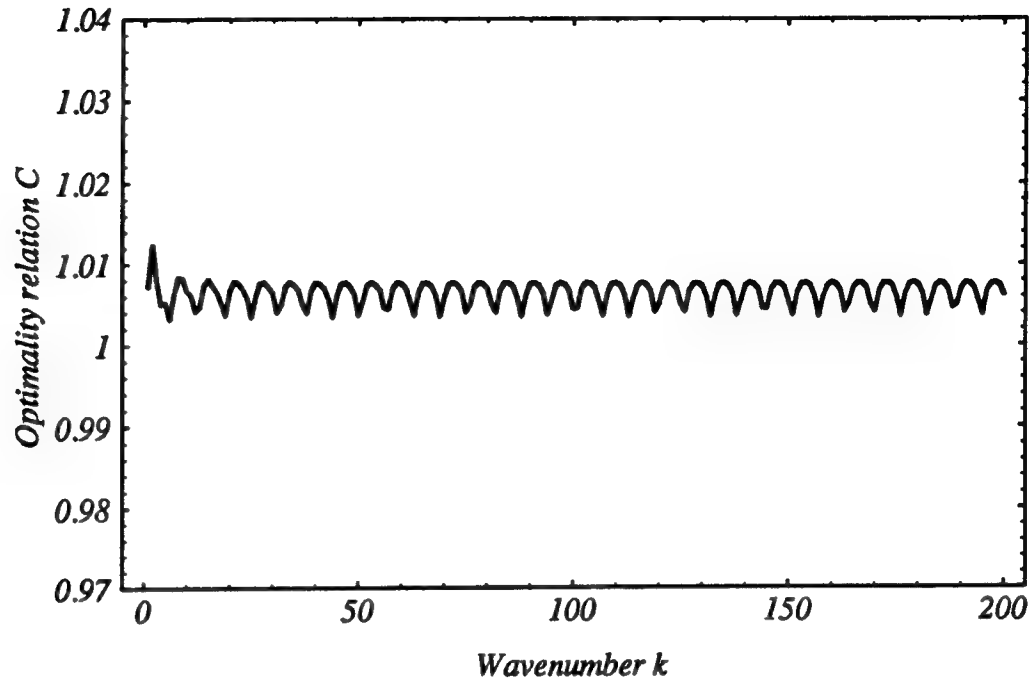


Figure 12: Relation  $e_{fe}/e_{opt}$  of the finite element error to the minimal error in the  $h - p$  approximation space in  $H^1$ -seminom for  $p = 1$  on square normalized mesh with  $k^2h = 1$

proposition holds:

**Theorem 3** Assume that the exact solution  $u$  is of regularity  $l + 1$ . Then, if  $hk \leq \alpha < \pi$ ,

$$|u - u_{fe}|_1 \leq C_1(p) \left( 1 + C_2 k \left( \frac{kh}{2p} \right) \right) \left( \frac{h}{2p} \right)^m |u|_{m+1} \quad (21)$$

with

$$C_1(p) = \left( \frac{e}{2} \right)^p (\pi p)^{-1/4}, \quad (22)$$

$$m = \min(l, p)$$

and  $C_2$  not depending on  $h, k$  and  $p$ .

This theorem generalizes the quasioptimal estimate (13) of Theorem 2. Indeed, if  $k^2h/2p$  is bounded then the expression in the brackets is constant, yielding eq (14). Furthermore, for oscillatory solutions

$$\tilde{e} \leq C_1(p) \left( \frac{hk}{2p} \right)^m + C_2(p) k \left( \frac{hk}{2p} \right)^{m+1} \quad (23)$$

follows. Comparing this estimate to estimate (20), one is led to the following definition.



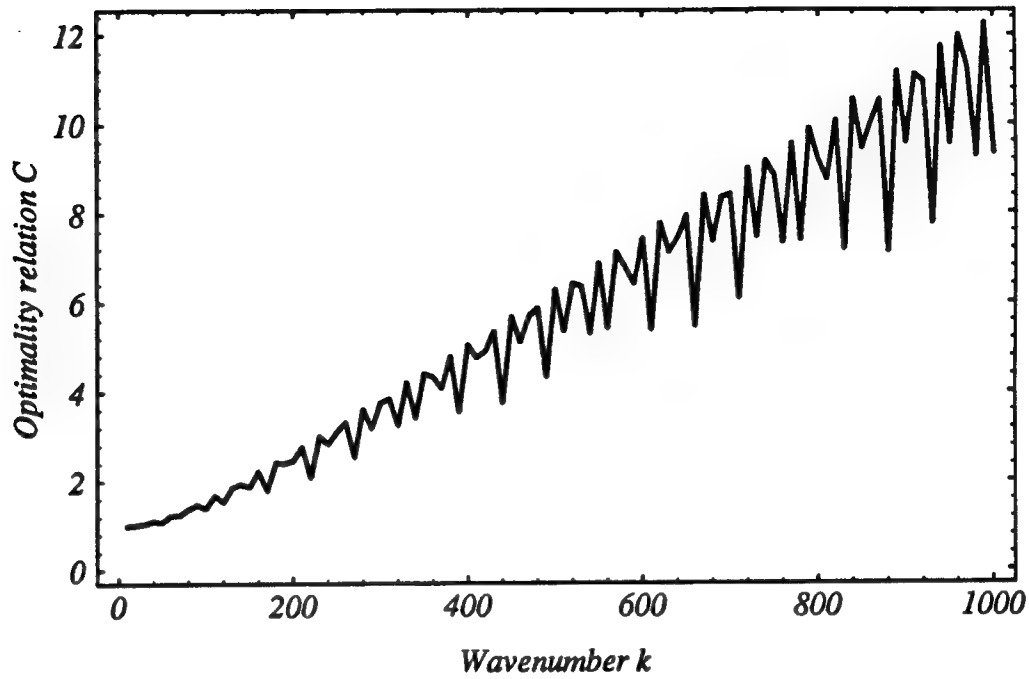


Figure 13: Relation  $e_{fe}/e_{opt}$  of the finite element error to the minimal error in the  $h - p$  approximation space in  $H^1$ -seminom for  $p = 1$  on normalized mesh with  $hk = 0.1$ .

**Definition 3 (Pollution):** Consider a Helmholtz problem with parameter  $k$ , given in variational formulation on a normed space  $V$  with norm  $\|\cdot\|_V$ . Assume unique existence of an exact solution  $u \neq 0 \in V$  and a finite element solution  $u_{fe} \in V_h \subset V$ . Assume further that an estimate of the form

$$\tilde{e}_{fe} := \frac{\|u - u_{fe}\|_V}{\|u\|_V} \leq C(k) \inf_{v \in V_h} \frac{\|u - v\|}{\|u\|} \quad (24)$$

holds.

Then, if  $C$  can be written in the form

$$C(k) = C_1 + C_2 k^\beta \theta^\gamma \quad (25)$$

where

- $\beta > 0$ ,  $\gamma \geq 0$ ;
- $\theta = hk/p$  is the scale of the  $h$ - $p$ -mesh and
- $C_1, C_2$  are independent of  $h, k$  and  $p$ ,

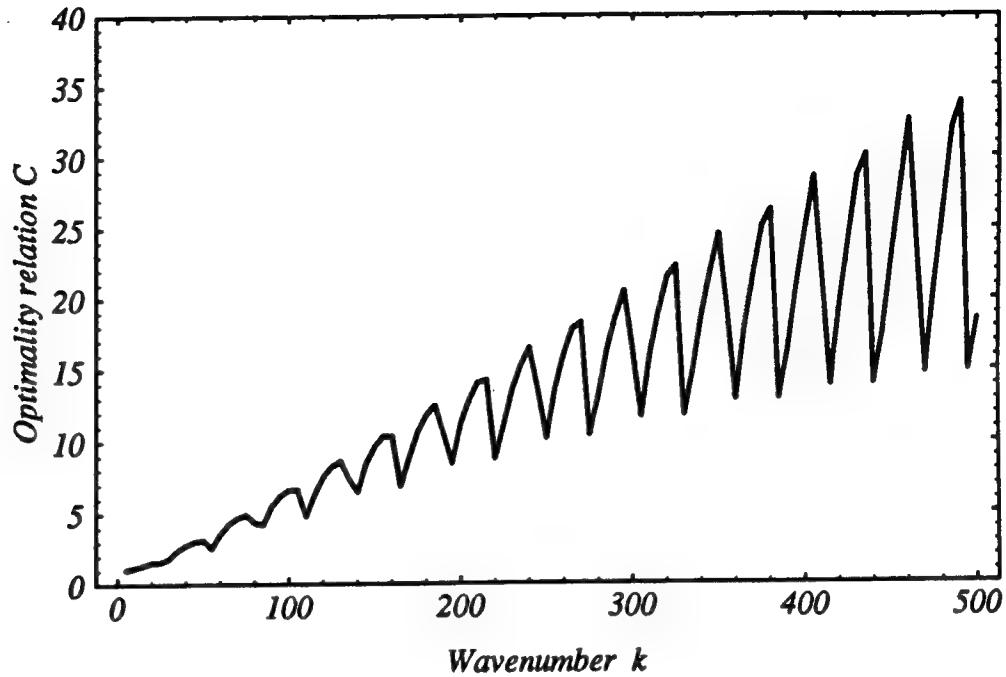


Figure 14: Relation  $e_{fe}/e_{opt}$  of the finite element error to the minimal error in the  $h - p$  approximation space in  $H^1$ -seminorm for  $p = 4$  on  $h - p$  normalized mesh with  $hk/2p = 0.5$ .

the finite element solution is said to be polluted and the term  $C_2 k^\gamma$  is called pollution term.

**Remark 7:** This definition slightly modifies definition 1.2 in [6] to accomodate the context of this investigation. Note that the pollution term here includes the error of best approximation as a factor. It is further assumed that the estimate is known to be sharp - either by theoretical proof or by numerical evaluation.

**Remark 8:** Speaking in terms of the previous paragraph, a polluted estimate is of dispersive character, whereas a nonpolluted estimate is nondispersive. Hence, any modified finite element method reducing or eliminating the pollution effect equivalently reduces or eliminates the dispersive effect of the numerical solution.

For numerical evaluation, consider again the exterior Helmholtz problem with Dirichlet condition at  $x = 0$  (example 1).

In the case of piecewise linear approximation ( $p = 1$ ), one obtains from (23) (cf. also [14], Theorem 5):

$$\tilde{e} \leq C_1(hk) + C_2 k(hk)^2$$

where  $C_1, C_2$  do not depend on  $h, k$  ( $p$  is fixed).

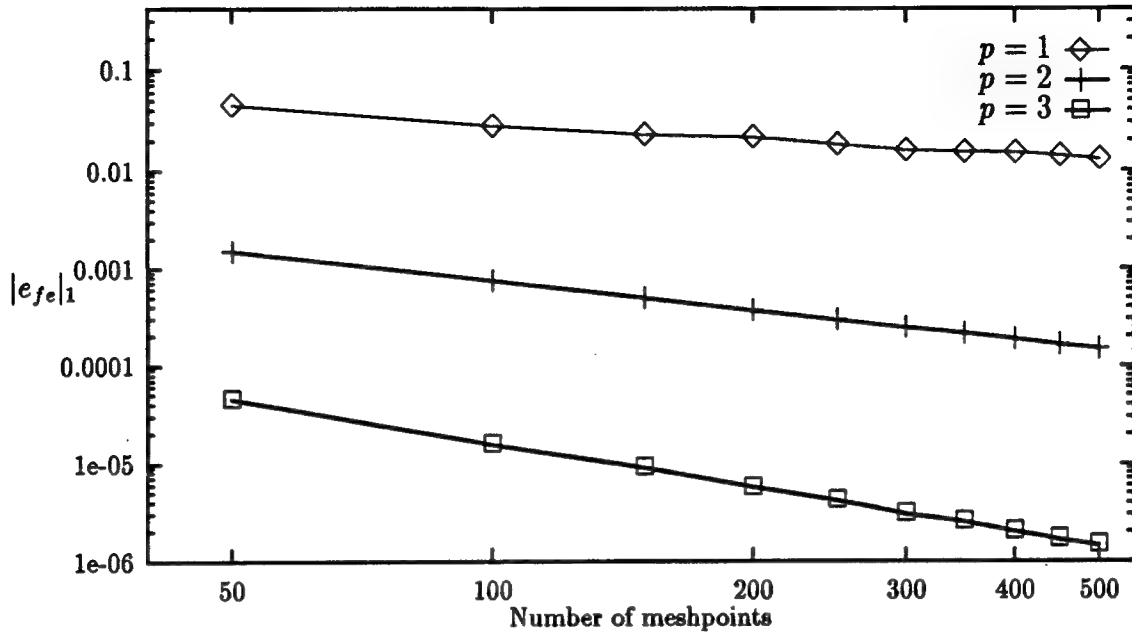


Figure 15: Error on square-normalized mesh for different  $k$  and  $p$  with  $k = (pn)^{1/2}$

A typical error plot in log-log-scale, both for the error of the interpolant (best approximation) and the error of the FE-solution, is shown in Fig. 16.

For a more detailed discussion, three critical points are displayed on the abscissa on this plot, marking four ranges of error behaviour.

1.  $n \leq n_o$ : The point  $n = n_o$  is the limit of resolution given by the relation  $hk = \pi$  (two elements per wavelength). For  $n \leq n_o$ , the error of the interpolant is 100%. Since  $\pi < K_o = \sqrt{12}$ , the finite element solution is a decaying wave – cf. paragraph on cutoff frequencies. Hence, in the first range marked on the plot, below the limit of resolution, both the finite element solution and the interpolant do not converge to the exact solution at all.
2.  $n_o < n \leq N_o$ : The point  $n = N_o$  is the critical number of DOF for the finite element solution, i.e. the minimal meshsize for which the error is (and stays, if the mesh is refined) below 100%. In [14] it is shown by physical argument and numerical evaluation that

$$N_o = \sqrt{\frac{k^3}{24}}$$

can be used as an approximate formula to compute  $N_o$ .

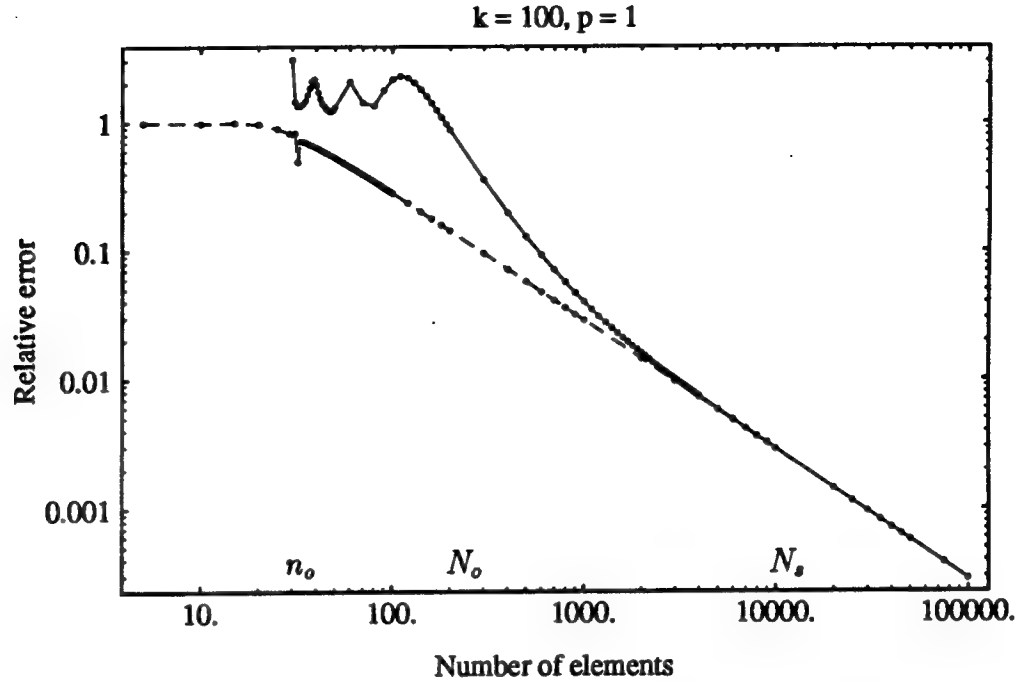


Figure 16: Relative errors of the interpolant and the finite element solution in  $H^1$ -seminorm for  $p = 1$  and  $k = 100$

For  $n_o < n \leq N_o$ , the error of the interpolant goes down exponentially with rate  $-1$  as  $h$  decreases, whereas the error of the finite element solution oscillates with amplitudes of more than 100%. Note that, in accord with the stability theorem, the finite element solution exists and is uniquely defined, but the numerical wave can by no means be considered a reliable approximation of the exact wave. In other words, in the first range following the limit of resolution, the interpolant converges with optimal rate, but the finite element solution is not converging.

Referring to the error estimate (23), the pollution term is significantly larger than the approximation term in the range under consideration. So, for the given example, the error of interpolation is  $\approx 15\%$  whereas the finite element error still amounts to  $\approx 100\%$ .

3.  $N_o < n \leq N_s$ : The point  $n = N_s$  is the point on the abscissa to the right of which the finite element solution shows quasiotimal, nondispersive convergence behaviour. As a consequence of Theorem 2, the formula to compute  $N_s$  is

$$N_s = \frac{1}{c^*} k^2$$

where, theoretically,  $c^*$  is a small constant. Following the observation of Remark 4,  $c^* = 1$  has been assumed in the computations.

In the interval under consideration,  $N_0 < n \leq N_s$ , the best approximation further converges with optimal rate while the convergence behaviour of the finite element solution is still governed by the pollution term. Note that the prevalence of this term leads to a *superoptimal* rate of convergence, namely, exponential decay  $n^{-2}$ , as compared to  $n^{-1}$  for the best approximation. Thus, the error of the finite element solution visibly goes down towards the interpolation error until the pollution term becomes negligible at  $n \approx N_s$  (i.e. when  $k^2 h$  is sufficiently small).

This third region is actually the region of practical interest. The finite element error is sufficiently small while the mesh is still reasonably coarse. In particular, the error is within this region if, using the estimate (23) or, equivalently, the result from dispersion analysis on the size of the phase difference, one constrains the magnitude of  $k^3 h^2$ . With a constraint of this form it is guaranteed that the numerical wave stays sufficiently close to the exact wave, i.e. is a reliable solution, for all  $k$  – see Fig. 17.

4.  $n > N_s$ : This is the range where both the interpolant and the finite element solution converge with optimal rate  $n^{-1}$ . In particular, for the considered model problem  $C_{opt} \rightarrow 1$  as  $h \rightarrow 0$  with the visual effect that the plotted errors coincide to the right of  $N_s$  within the resolution accuracy of the plot. Both the amplitude and the phase error of the finite element solution are very small, to the expense, however, of an overrefined mesh. The stepwidth  $h$  has to be chosen by square-normalisation w.r. to  $k$ .

**Remark 9:** It is understood that the boundaries of the ranges commented on above are not defined precisely as points, but rather in a fuzzy sense.

While all four intervals outlined here are present for any value of  $k$ , their length depends significantly on  $k$ . In particular, for low wavenumber, as mentioned before, the finite element solution is close to the interpolant throughout the range of convergence, hence practically only the range 4. occurs in the error plot. Fig. 18. shows that for  $k = \pi$  the finite element error is close to the minimal error throughout the region of convergence: though slightly different in magnitude, the errors are not distinguishable in the plot. The phase lead in this case is not an essential source of numerical error. On the other hand, for large  $k$  the absolute phase lead of order  $k^3 h^2$  is of prevailing influence; so, for  $k = 100$ , the finite element error is greater than 100% even if, following the common rule, ten linear elements are used to resolve one wavelength – as can be seen from Fig. 16.

Returning now to the estimate (23) it should be emphasized that the first term reflects the error of interpolation (the amplitude error inside the elements, as determined by the choice of approximating functions) whereas the second term reflects the phase error. The estimate shows that the error of the finite element solution on normalized mesh for large

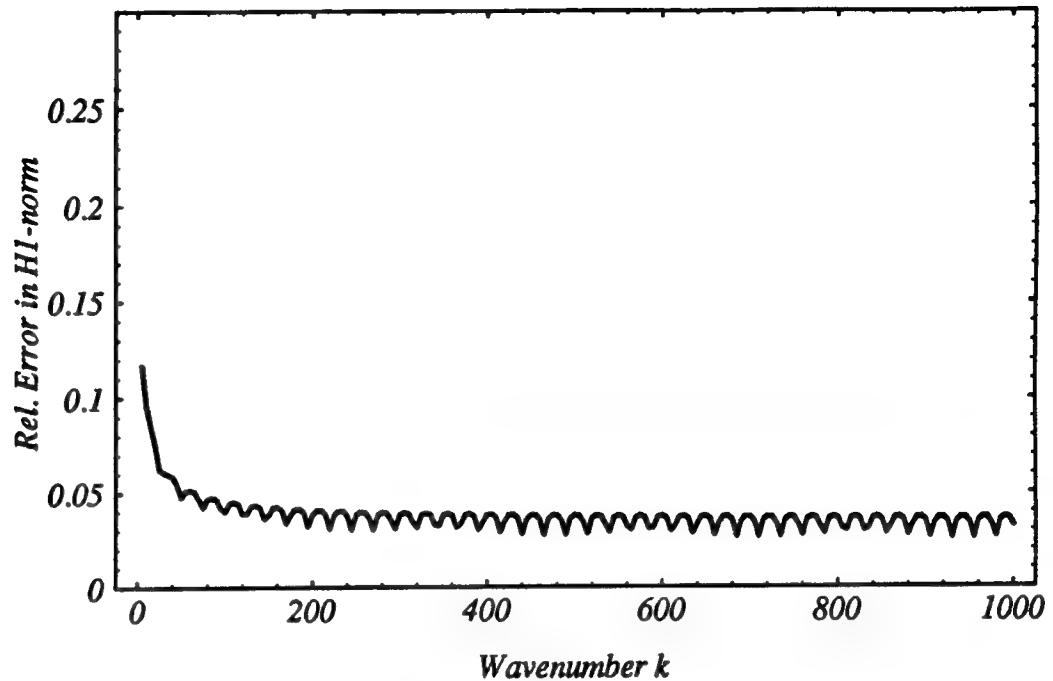


Figure 17: Relative error of the finite element solution in  $H^1$ -seminorm for  $p = 1$  with constraint  $h^2 k^3 = 1$  for  $k = 1, 1000, 1$ .

wavenumber is essentially equal to the phase error. The phase error in integral norm is, in turn, of the same order as the (absolute, not the normalized) phase difference as investigated in dispersion analysis. Hence, if one uses the conclusions of dispersion analysis to control or minimize the phase lead one equivalently influences the integral error in the preasymptotic range, and vice versa.

On the other hand, the finite element error on square-normalized mesh is essentially equal to the error of interpolation; the finite element solution is then "almost" in phase with the exact solution.

**Remark 10:** The estimate (23) has been shown with the assumption of uniform mesh. However, numerical experiments on extremely nonuniform meshes (with  $p = 1$ ) have shown the same error behaviour [6].

**Example 2: Exterior problem with Robin conditions.** In this second example, the ordinary Helmholtz equation is considered on  $\Omega = (0, 1)$  with Robin boundary conditions both at  $x = 0$  and  $x = 1$ . The boundary value problem is then equivalent to the



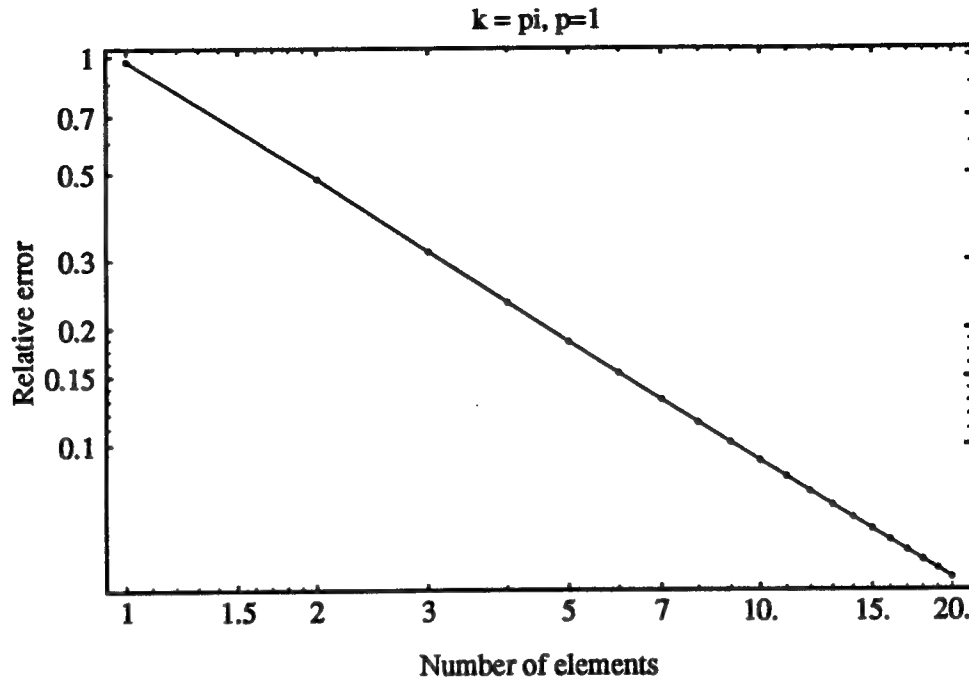


Figure 18: Relative errors of the interpolant and the finite element solution in  $H^1$ -seminorm for  $p = 1$  and  $k = \pi$

variational problem: Find  $u \in H^1(\Omega)$  such that

$$B^*(u, v) = (u', v') - k^2(u, v) - ik\langle u, v \rangle = (f, v), \quad (26)$$

where

$$\langle u, v \rangle = u(0)\bar{v}(0) + u(1)\bar{v}(1),$$

holds for all  $v \in H^1(\Omega)$ .

This problem has been analyzed by Douglas et al. [10]. Demkowicz [9] has shown that, up to additional material parameters, the problem (26) is equivalent to a one dimensional fluid structure interaction model. For a detailed analysis of Galerkin methods for one-dimensional fluid structure interaction problems, including further numerical evaluation, see [4].

For numerical evaluation, let again  $f \equiv 1$ . Due to the absence of Dirichlet boundaries, the  $H^1$ -seminorm is not equivalent to the  $H^1$ -norm; the results are therefore measured in the norm

$$\|u\|_* = \left( \|u'\|^2 + |u(0)|^2 \right)^{1/2}$$

which will be referred to as the  $H^1$ -norm in the following.

In Fig. 19, the error of the finite element solution in  $H^1$ -norm is plotted and compared to the error of the best approximation in this norm. The same principal behaviour is

observed as for the Dirichlet-fixed case. However, the optimality constant is not going to 1 as  $h$  goes to zero. The critical points  $N_o$  (error of fe-solution stays below 100%) are at about the same positions in both cases. The plots show that the optimality constant tends to a magnitude of  $\approx 3.5$  independently of  $k$ . This observation is in accord with the estimate (14), establishing nondispersive convergence behaviour on square-normalized mesh.

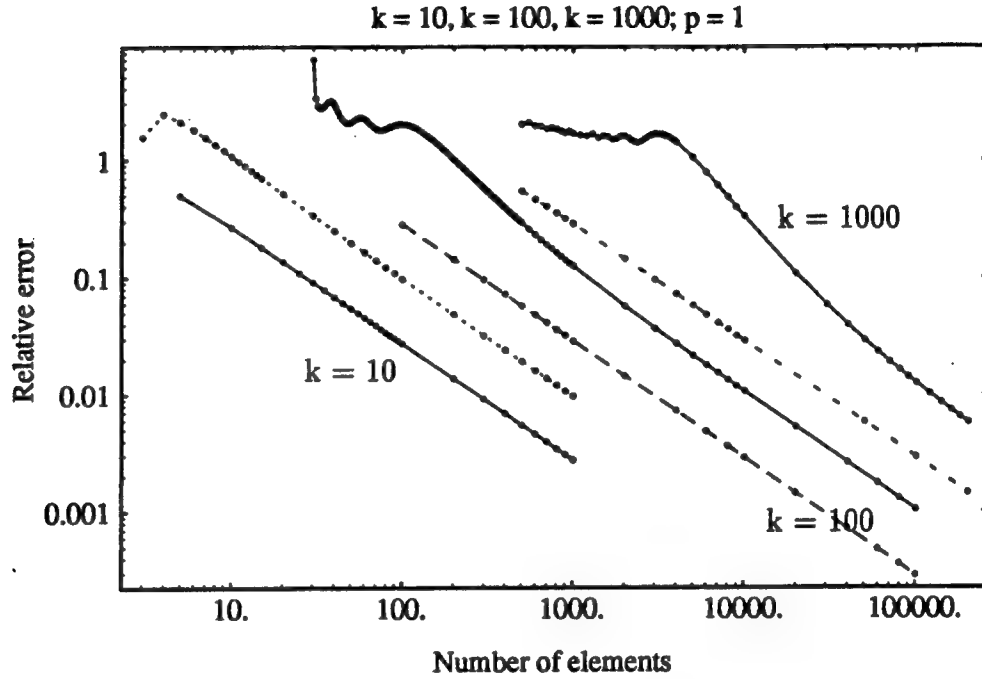


Figure 19: Exterior problem with Robin conditions: Relative errors of the interpolant and the finite element solution in  $H^1$ -norm for  $p = 1$  and  $k = 10, 100, 1000$

**Error analysis on normalized mesh II:** Turn now to the general estimate for  $p > 1$ . Comparing eq (23) to the estimation of the phase difference in Theorem 1 one observes that the pollution term is, unlike in the case  $p = 1$ , not of the order of the phase lead if  $p > 1$ . By more refined analysis, employing specific properties of the approximating polynomials that allow estimation by dual Sobolev norms, one shows that the order of  $hk/2p$  in the pollution term can be further raised for  $p > 1$  yielding the estimate ([15], Theorem 3.7 and Corollary 3.2):

$$\tilde{e} \leq C_1(p) \left[ \left( \frac{hk}{2p} \right)^m + C_1(p)C_2(p)k \left( \frac{hk}{2p} \right)^{p+m} \right] \quad (27)$$

with  $m = \min(l, p)$  where  $l + 1$  is the regularity of the solution  $u$ . Assuming maximal regularity ( $l \geq p$ ), the estimate becomes

$$\tilde{e} \leq C_1(p) \left[ \left( \frac{hk}{2p} \right)^p + C_1(p)C_2(p)k \left( \frac{hk}{2p} \right)^{2p} \right] \quad (28)$$

In this estimate, the pollution term is of the order of the phase difference and the constant  $C_1(p)$  is growing with  $p$  by the rate of eq (22), Theorem 3. The second constant is theoretically of order

$$C_2(p) \approx 2^p p!.$$

Hence, unlike in eq (23) and in the estimate for the phase lead, it could not be excluded in the proof of the above mentioned theorem that the constant  $C_2$  may grow with  $p$ . Since no numerical evidence has yet been found for this growth of the pollution term with  $p$  the theoretical estimate of this term is not proven to be sharp – see also the computational results for the model problem below. On the other hand, though the growth rate is significant by itself it has, even if present, relatively little influence on the practical conclusions from estimates (27), (28) as will be seen below.

Conclusions from the estimate (28) are:

*First*, as in the case  $p = 1$ , dispersion analysis of the (absolute) phase difference and numerical analysis of error norms (on normalized mesh) are equivalent w.r. to the reliability of the finite element solution.

*Second*, writing

$$\tilde{e} \leq C_1 \left( \frac{hk}{2p} \right)^p \left[ 1 + C_1 C_2 k \left( \frac{hk}{2p} \right)^p \right] \quad (29)$$

one concludes that the error is in the quasioptimal range provided

$$k \left( \frac{hk}{2p} \right)^p \leq c^* \quad (30)$$

where  $c^*$  is sufficiently small. Obviously this is, for higher  $p$ , a much weaker condition than the assumption  $k^2 h < c^*$  of Theorem 1.

For numerical evaluation, consider again the exterior problem with constant data. In Table 2, numerical values are compared for the critical number  $N_c$  (inset of quasioptimal behaviour) computed for  $c^* = 0.5$ :

- as computed from the assumption on  $h - p$  square-normalized mesh in Theorem 1 and
- as computed from the relation (30), following from the preasymptotic estimate II.

**Table 2:** Critical numbers  $N_s$  for the range of asymptotic convergence of the finite element solution: prediction from preasymptotic estimate I vs. prediction from preasymptotic estimate II with  $c^* = 0.5$ .

$p$	1	2	3	4	5	6
$N_s, I$	2500	1240	834	625	500	417
$N_s, II$	2500	125	59	20	13	9

Hence, if for higher  $p$  not the assumption of square-normalized mesh but the weaker condition (30) determines the critical number  $N_s$ , the range of asymptotic convergence is *significantly enlarged* with growing  $p$ . As the table indicates, one gains most significantly when passing from  $p = 1$  to  $p = 2$  or  $p = 3$ . The numbers given for  $N_s, II$  in Table 2 are in accord with numerical results as presented in Fig. 20

The "bumps" in the error lines in Fig. 20 occur at the meshsizes for which  $kh = l\pi, l = 1, 2, \dots$ . On these particular meshes, the approximation problem is locally reduced to approximating the function  $\sin x$  or  $\cos x$ , resp. In this case, raising the order of polynomial approximation by one does not in general decrease the error of approximation – cf. [15].

Finally, it is demonstrated on the model problem that the estimate (27) yields a  $p$ -adaptive rule for error control by appropriate choice of the meshsize. Assuming that the error is dominated by the pollution term, put

$$\varepsilon = C_1^2(p)C_2(p) \left( \frac{k}{2np} \right)^{2p} \quad (31)$$

where  $\varepsilon$  is an a priori given tolerance. The number of DOF is then

$$np = \frac{k}{2} \left( \frac{C_1^2 C_2 k}{\varepsilon} \right)^{1/2p}. \quad (32)$$

In Table 3, the number of DOF needed to achieve accuracy of 10, 20 or 40%, resp., is tabulated where measured magnitudes are compared to the estimated number in eq (32). The results are given for wavenumbers  $k = 10$  and  $k = 100$ . The estimated number has been computed putting  $C_1 = C_2 \equiv 1$ , i.e. not considering the constants of the estimation. It can be seen that, for the model problem, safe upper estimates for the appropriate meshsize are obtained by controlling the order of the pollution term or, equivalently, the order of the phase lead of the Galerkin finite element solution.

**Table 3:** Controlling the error in integral norm by controlling the order of the phase lead: results for  $k = 50, k = 100, p = 1 \dots 6$  and  $\varepsilon = 0.1, 0.2, 0.4$ .

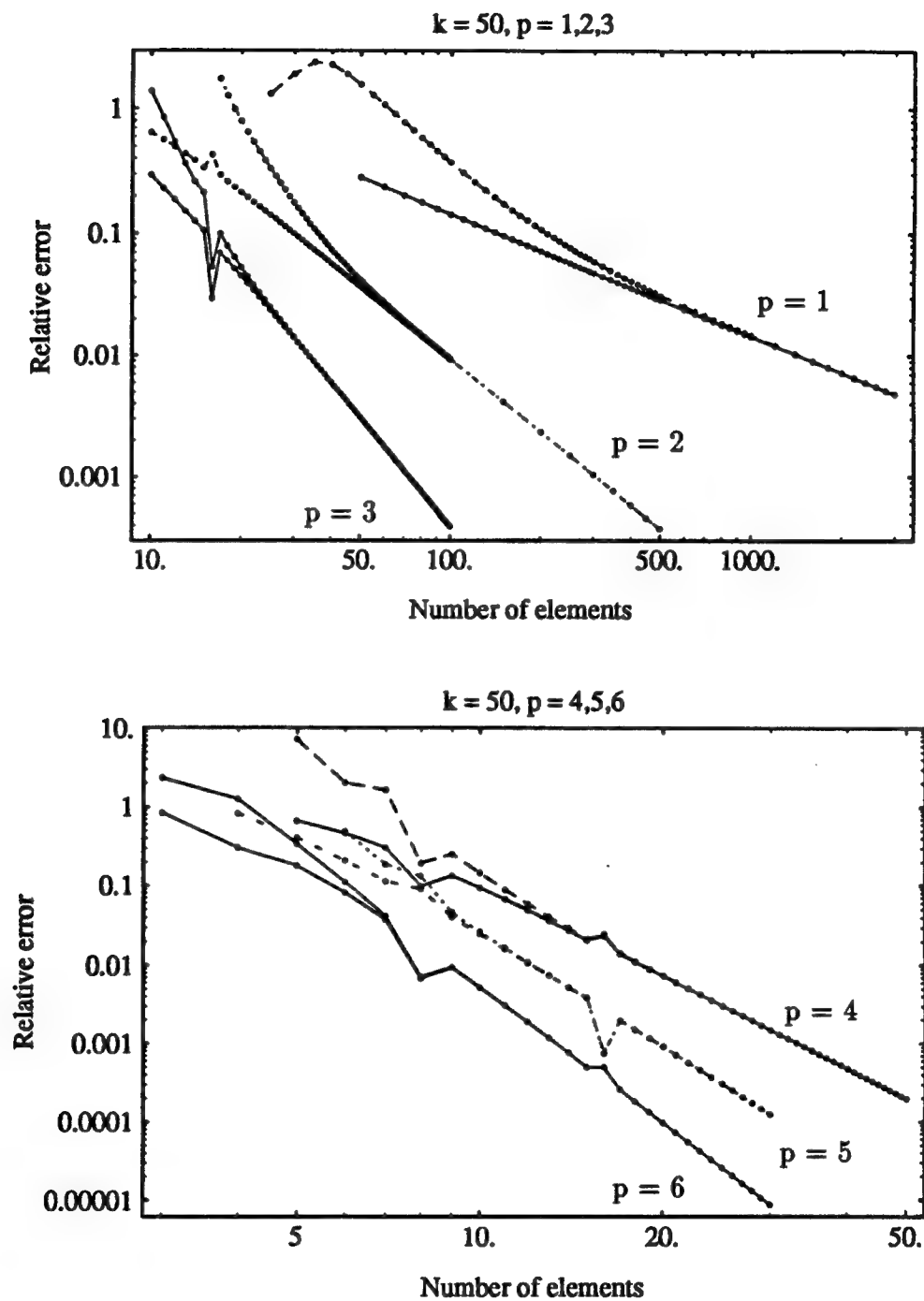


Figure 20: Error of finite element solution versus error of best approximation for  $k = 50$  and  $p = 1, 2, 3$ ;  $p = 4, 5, 6$ .

$p$	1	2	3	4	5	6	Tolerance $\varepsilon$	wavenumber $k$
# DOF, Fig.	210	72	48	44	40	36	0.1	50
# DOF, eq.	560	118	70	54	47	42		
# DOF, Fig.	150	56	45	40	30	30	0.2	
# DOF, eq.	395	99	63	50	43	40		
# DOF, Fig.	92	48	39	36	30	30	0.4	
# DOF, eq.	280	86	56	46	41	37		
# DOF, Fig.	580	170	120	92	85	78	0.1	100
# DOF, eq.	1581	281	158	119	100	89		
# DOF, Fig.	400	140	105	84	75	72	0.2	
# DOF, eq.	1118	236	141	109	93	84		
# DOF, Fig.	280	116	87	76	70	64	0.4	
# DOF, eq.	790	199	125	100	87	79		

Graphical illustration for the table is given in Figs. 20, a,b (for  $k = 50$ ) and in Figs. 21, a,b for  $k = 100$ . In the latter figures, the relative error of the finite element solution is plotted against the number of meshpoints (a) and against the number of DOF  $np$  (b). The advantage of the higher  $p$ -versions is obvious.

Also when taking into account the theoretically established growth of the pollution term, including the constants  $C_1, C_2$  with the approximation order  $p$ , one obtains upper estimates for the appropriate meshsize that go down with  $p$ . This can be seen from Table 4, where upper estimates for the number of DOF are listed and compared to the measured number of DOF.

Table 4: Controlling the error in integral norm by controlling the pollution term: results for for  $k = 50$ ,  $p = 1 \dots p = 6$  and  $\varepsilon = 0.1$ .

$p$	1	2	3	4	5	6
# DOF, Fig.	210	72	48	44	40	36
# DOF, eq.	807	215	151	133	126	123

**Remark 11:** If the regularity of the solution is smaller than  $p + 1$  - this is, e.g. the case when higher approximation is used for problems with unsmooth, like Dirac or piecewise constant, data - the estimate (27) applies. The pollution term is then  $C_2 k \theta^{p+m}$  with  $m < p$ . It's order is lower than the order of the phase difference, but eq (30) still determines the range of quasioptimal convergence.

Taking out the best approximation term, one has

$$\tilde{e} \leq C_1 \theta^m [1 + C_1 C_2 k \theta^p].$$

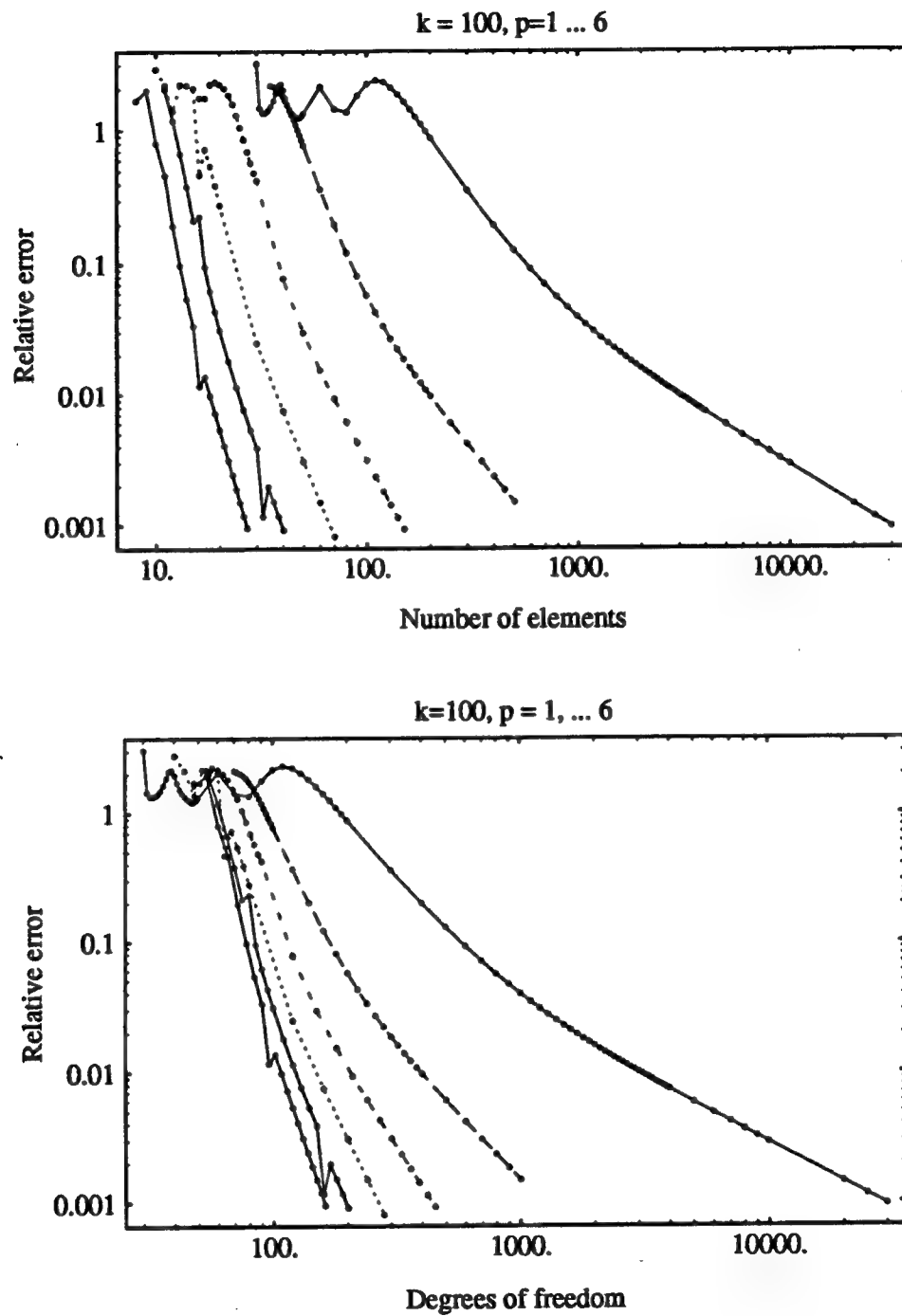


Figure 21: Error of finite element solution as a function of the meshsize  $n$  (a) and as a function of the number of DOF (b) for  $k = 100$  and  $p = 1, 2, \dots, 6$ .

The term in the brackets depends only on the order of approximation but not on the regularity of the solution. Consequently, for low regularity and high approximation, the pollution term essentially does not influence the numerical error.

**Remark 12:** The phase accuracy or, equivalently, the pollution term in the finite element solution, can be significantly reduced for one-dimensional problems using suitable modification of the Galerkin FEM. While the improvement may be spectacular in one-dimensional problems for  $p = 1$ , the improvement becomes less significant as  $p$  is raised to higher order of approximation.

For two- or three-dimensional problems, where the pollution term cannot be eliminated from the error in general (or, alternatively, phase accuracy in all directions of propagation is not simultaneously achievable by the same modification – see [6], [20]), the relative improvement of the solution quality is still less significant compared to the one-dimensional case. On the other hand, multidimensional acoustic computations, specifically those with medium or high wavenumbers involved, are extremely costly, hence even slight reductions of the size of the numerical model may be welcome in practical application.

**Error estimation in  $L^2$ -norm:** So far, all integral estimates in this section have been given in the  $H^1$ -seminorm, i.e. the square integral  $L^2$ -norm of the derivatives of the solution. For the model problem under consideration, this norm is equivalent to the  $L^2$ -norm of the solution itself. However, the rates of convergence in both norms are different. The following exemplary discussion is given for the case of piecewise linear approximation ( $p = 1$ ).

Douglas et al. showed that for  $p = 1$  and  $k^2 h$  sufficiently small (i.e. on square-normalized mesh) the estimate ([10], Lemma 2.6):

$$\|u - u_{fe}\| \leq C(1 + k)^2 h^2 \|f\| \quad (33)$$

holds with a constant  $C$  not depending on  $h, k$ .

**Remark 13:** Bayliss et al. had stated the following proposition for higher approximation – also with the assumption of square normalized mesh ([7], eq (2.5)): Assuming that the approximation order of the finite element solution is  $p$  and the exact solution is at least  $m = p + 1$  times differentiable then

$$\|u - u_{fe}\| \leq C_m h^m (1 + k^{m+1}) \|u\| \quad (34)$$

holds, where  $C_m$  depends on  $m$ . Using the stability estimate  $\|u\| \leq C k^{-1} \|f\|$  it is easy to see that eq (33) follows from eq (34) for  $p = 1$ .

On normalized mesh, the following proposition holds.



**Theorem 4** Let  $u \in H^{(1)}(\Omega)$ ,  $u(0) = 0$  and  $u_{fe} \in S_h^1(\Omega) \subset H^{(1)}(\Omega)$  be the solution and the finite element solution to the VP (6), respectively. Assume that  $hk < 1$ .

Then for  $e := u - u_{fe}$

$$\|e\| \leq (1 + Ck) \left(\frac{h}{\pi}\right)^2 |u|_2 \quad (35)$$

holds, where  $C$  does not depend on  $h, k$  and  $p$ .

The proof of this theorem is similar to the proof of the  $H^1$ -estimate ([14], Theorem 5). Namely, writing  $u - u_{fe} = u - u_I + u_I - u_{fe}$ , one has by triangular inequality

$$\|u - u_{fe}\| \leq \|u - u_I\| + \|z\|$$

where  $z := u_I - u_{fe}$  can be shown to be the solution of the variational problem

$$\forall v \in V_h : \quad B(z, v) = k^2(u - u_I, v).$$

From the discrete Green's function representation of  $z$  it is then concluded that

$$\|z\| \leq C_1 k \|u - u_I\|$$

with  $C$  independent of  $h, k$ .

Thus  $\|u - u_{fe}\| \leq (1 + C_1 k) \|u - u_I\|$ , and the statement follows from a standard approximation result. Details can be found in [14].

For a first numerical evaluation, consider in Fig. 22 the convergence of the relative error in  $H^1$ -seminorm compared to the  $l^2$ -vectornorm for  $k = 100$ .

One observes:

- In both norms, the error oscillates with magnitudes of  $> 100\%$  on coarse mesh before reaching the critical point  $N_o$  ("knee" in the plots after which the errors stay below  $100\%$ ) at about the same number  $n$ .
- After entering the range of convergence, both errors initially decrease with the same rate  $n^{-2}$ . Then, while the error in  $l^2$ -norm continues to converge with that very rate, the error in  $H^1$ -norm deviates to assume it's optimal rate  $n^{-1}$  for large  $n$ .

For an illustration of the dispersive character of the constant  $C(k) = 1 + C_1 k$  in eq (35), see Fig. 23. The error for constant scale, computed on a sequence of different magnitudes  $h, k$  on meshes scaled by  $hk = 1$ , is plotted.

Concluding this paragraph it is emphasised that, unlike the  $H^1$ -estimate, the  $L^2$ -estimate is of dispersive character throughout the range of convergence. On the other hand, the solution converges with optimal rate in the whole range.

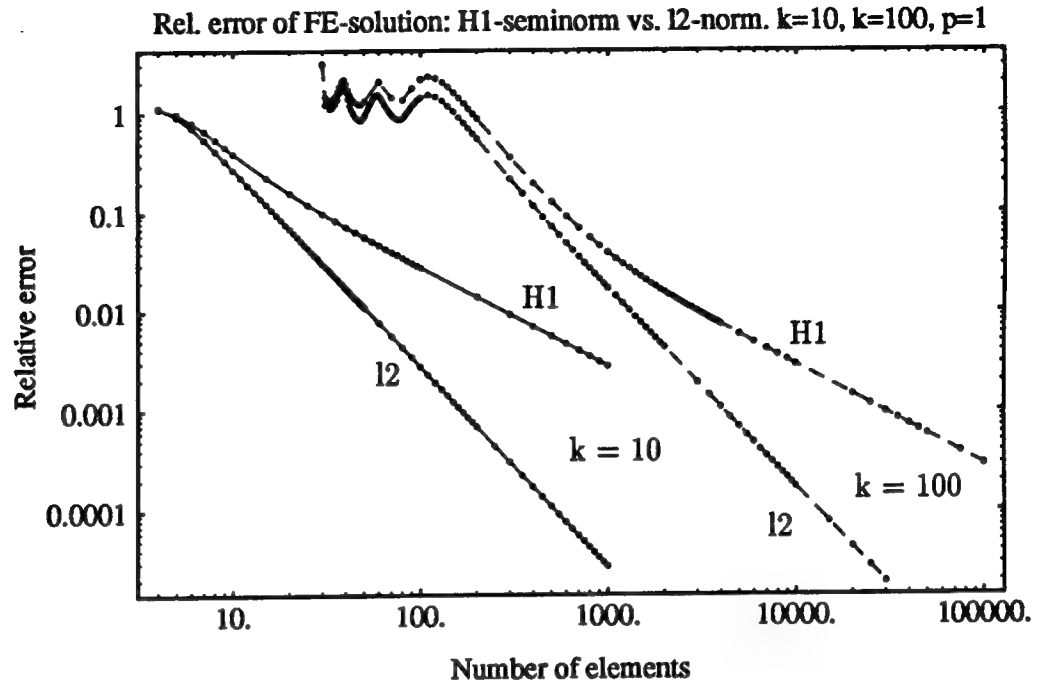


Figure 22: Relative errors finite element solution: results in  $H^1$ -seminorm compared to  $l_2$ -vectornorm for  $k = 10$  and  $k = 100$ ,  $p = 1$

**Singular values of condensation:** For general  $p$ , the determining equation for the discrete wavenumber  $k'$  is eq (8):

$$\cos(k'h) = -\frac{S_p(K)}{t_p(K)}$$

where  $S_p, T_p$  are rational polynomial functions of  $K = kh$  and, by Theorem 1,

$$k' - k = k \mathcal{O}\left(\frac{kh}{2p}\right).$$

In the theoretical investigation it was assumed that  $kh \leq \alpha < \pi$  to guarantee that the condensation is well defined. On the other hand, it was visible in the numerical tests for higher  $p$  that acceptable solutions are obtained if more than one half-wave is resolved by just one element of polynomial approximation. In dispersion analysis it was shown by Taylor expansion that the cutoff frequency also increases with growing  $p$ .

Consequently, the meshsize of the FE-approximation may extend the size of the halfwave for  $p \geq 2$  provided

- no condensation is applied or

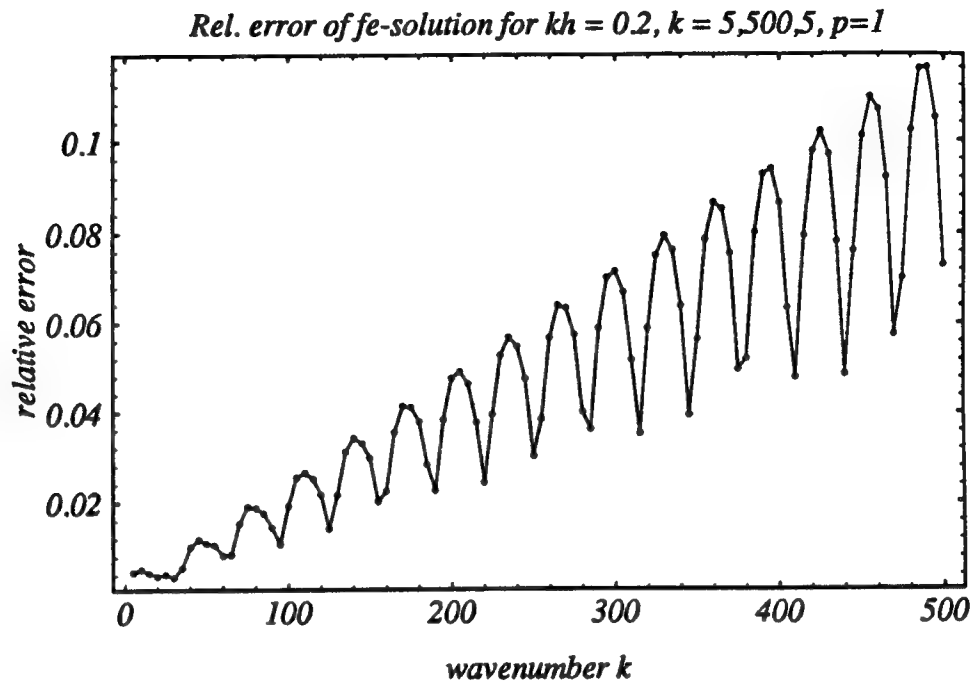


Figure 23: Relative error the finite element solution in  $l_2$ -vectornorm for  $p = 1$  on normalized mesh with  $kh = 0.2$ .

- condensation is applied but the meshsize  $h$  is such that  $K = kh$  is not close to any discrete eigenvalue of the condensation.

To find the discrete eigenvalues, one solves the EVP (in  $\lambda$ )

$$\begin{aligned} w'' + \lambda^2 w &= 0 \\ w(0) = w(1) &= 0 \end{aligned}$$

on the one-elemental approximation space

$$S_o^p = \{w \in S^p, w(0) = w(1) = 0\},$$

consisting of all polynomial bubble functions up to order  $p$ . The exact solutions are  $\lambda = \pi, 2\pi, \dots$

The discrete eigenvalues  $\lambda_1 \dots \lambda_{p-1}$  for different  $p$  are listed in table 5.

Table 5: Singular values  $K = \lambda_h$  of the local stiffness-mass-matrix and exact eigenvalues of the associated eigenvalue-problem for  $p = 2, \dots, 10$

$p/i$	1	2	3	4	5	6	7	8	9
2	3.16228	-	-	-	-	-	-	-	-
3	3.16228	6.48074	-	-	-	-	-	-	-
4	3.14612	6.48074	10.1060	-	-	-	-	-	-
5	3.14612	6.28503	10.1060	14.1597	-	-	-	-	-
6	3.14159	6.28503	9.44318	14.1597	18.7338	-	-	-	-
7	3.14159	6.28319	9.44318	12.6488	18.7338	23.8858	-	-	-
8	3.14159	6.28319	9.42494	12.6488	15.9505	23.8858	29.6460	-	-
9	3.14159	6.28319	9.42494	12.5681	15.9505	19.4030	29.6460	36.0290	-
10	3.14159	6.28319	9.42478	12.5681	15.7176	19.4030	23.0580	36.0290	43.041
$i * \pi$	3.14159	6.28319	9.42478	12.5664	15.7080	18.8496	21.9911	25.1327	28.274

**Two-dimensional results** Consider the homogeneous Helmholtz equation

$$\Delta u(x_1, x_2) + k^2 u(x_1, x_2) = 0 \quad (36)$$

on the unit square  $\Omega = (0, 1) \times (0, 1)$  with nonhomogeneous boundary conditions

$$iku + \frac{\partial u}{\partial n} = g_s \quad \text{on } \Gamma_s, \quad s = 1, 2, 3, 4 \quad (37)$$

where  $g$  is chosen such that the exact solution is

$$u = \exp(ikx)$$

with vector wavenumber  $k = (k_1, k_2)$  and  $|k| = k$ . The domain  $\Omega$  is covered with uniform quadratic mesh of stepsize  $h$ , and the finite element space is determined by the usual bilinear nodal shape functions. The Galerkin finite element solution  $u_{fe}$  is then computed from the linear system

$$(\mathbf{L} + k^2 \mathbf{M}) \mathbf{u}_h = \mathbf{r}_g \quad (38)$$

where  $\mathbf{L}$  is the stiffness matrix,  $\mathbf{M}$  is the mass matrix,  $\mathbf{r}$  is the discrete right hand side obtained from  $g$  and  $\mathbf{u}_h$  is the vector of nodal values of  $u_{fe}$  – see, e.g., [18]. It can be shown that the best approximations in  $H^1$ -seminorm and  $L^2$ -norm are easily computed from systems of the form

$$\tilde{\mathbf{L}} \mathbf{u}_{ba}^1 = \mathbf{r}_u^1$$

and

$$\tilde{\mathbf{M}} \mathbf{u}_{ba}^2 = \mathbf{r}_u^2$$

resp., where the right hand sides are computed from the exact solution  $u$  and the matrices on the left hand sides are, except for boundary stencils, identical with the stiffness or mass matrix, resp. The numerical results computed for this problem indicate that the Galerkin finite element solution in 2D behaves essentially like in the one-dimensional case.

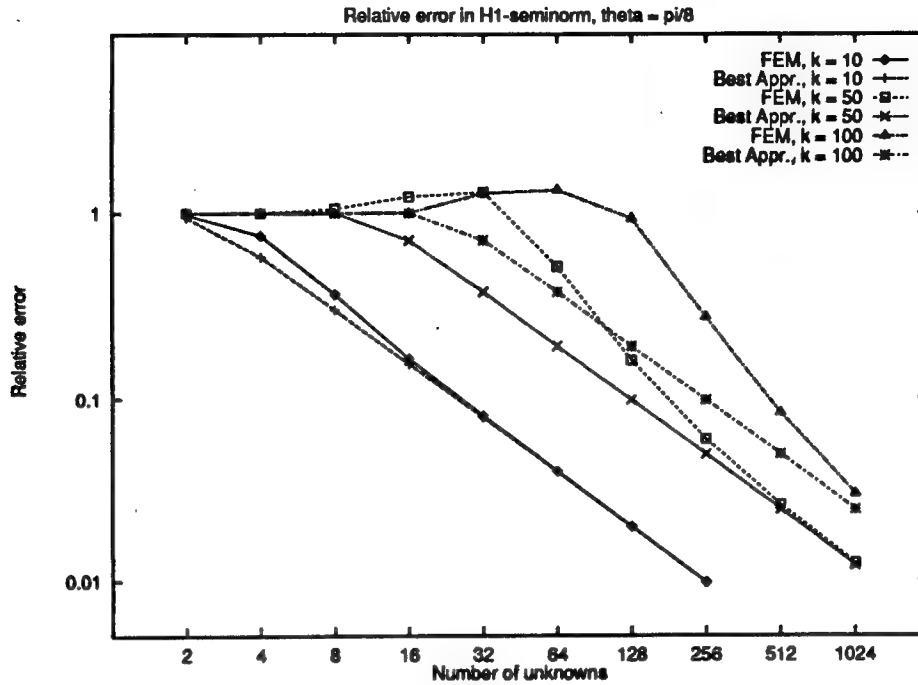


Figure 24: Relative errors for 2D Helmholtz problem: finite element solution vs. best approximation for  $k = 10$ ,  $k = 50$  and  $k = 100$

Consider in Fig. 24 the relative error of the finite element solution, compared to the error of best approximation, in  $H^1$ -seminorm for  $k = 10$ ,  $k = 50$  and  $k = 100$ . The vector components of the wavenumber  $\mathbf{k}$  are  $k_1 = k \cos \theta$ ,  $k_2 = k \sin \theta$ . The results plotted in the figure are computed for  $\theta = \pi/8$ . Comparing the results with the one-dimensional counterparts in Figs. 16 and 18 one observes similar behaviour of the numerical error.

In Fig. 25, the errors of the finite element solution and best approximation in  $H^1$ -norm and  $L^2$ -norm, resp., are shown for  $k = 50$ .

Using the computational results, one can check if the one-dimensional estimates principally hold also for the two-dimensional example. To this end, consider first for the  $L^2$ -norm the relation

$$\tilde{e}_{fe,2} := \frac{\|e\|}{\|u\|} \leq (1 + C_1 k) \frac{\|u - u_I\|}{\|u\|}$$

which was proven for the one-dimensional case (Theorem 4). Now, putting the two-dimensional  $\tilde{e}_{ba,2} := \|u - u_{ba}\|/\|u\|$  in the place of the one-dimensional  $\|u - u_I\|$ , one expects that the values  $C_1$  computed from

$$C_1 = \frac{1}{k} \left( \frac{\tilde{e}_{fe,2}}{\tilde{e}_{ba,2}} \right) \quad (39)$$

are constants that do not depend on  $h, k$ . The result of the computation is shown in Fig.

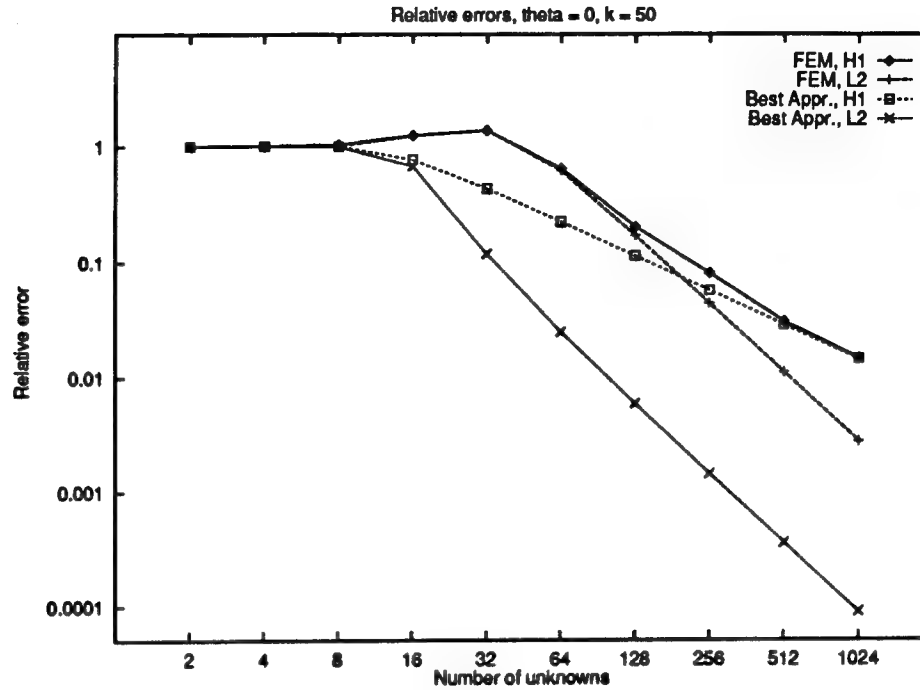


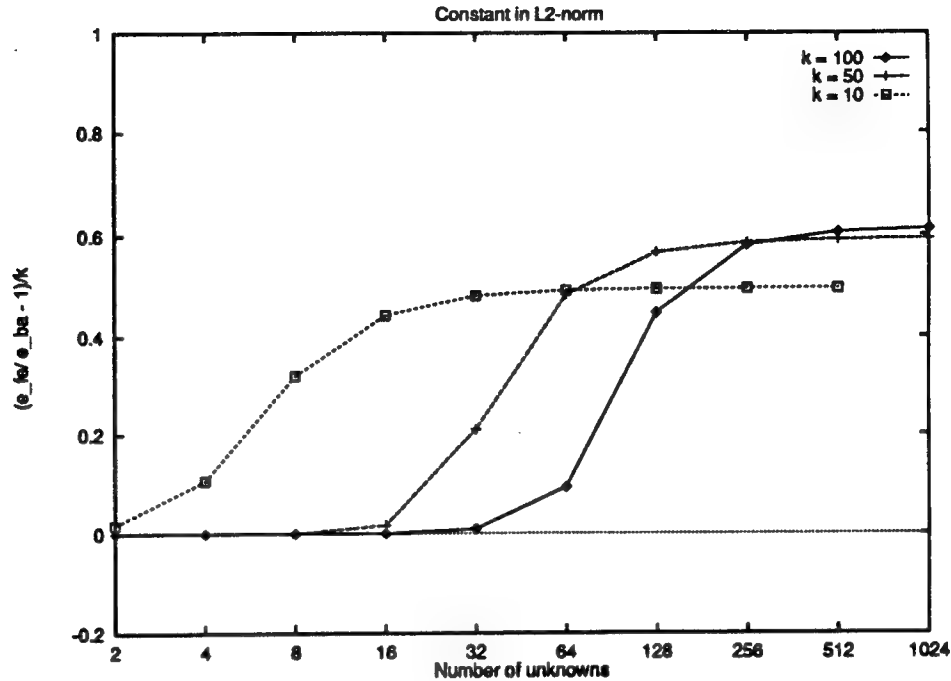
Figure 25: Relative errors for 2D Helmholtz problem: finite element solution vs. best approximation in  $L^2$ -norm and  $H^1$ -seminorm for  $k = 50$

26 and Table 6.

Table 6: Values  $C_1$  computed from (39)

$n/k$	100	50	10
2.	0.8003E-05	0.1706E-01	0.6486E-04
4.	0.1421E-04	0.1064E+00	0.1191E-03
8.	0.2967E-04	0.3208E+00	0.2997E-03
16.	0.7543E-04	0.4424E+00	0.1648E-01
32.	0.8667E-02	0.4812E+00	0.2103E+00
64.	0.9278E-01	0.4916E+00	0.4829E+00
128.	0.4471E+00	0.4943E+00	0.5662E+00
256.	0.5808E+00	0.4950E+00	0.5862E+00
512.	0.6059E+00	0.4951E+00	0.5911E+00
1024.	0.6114E+00	-	0.5923E+00

A more detailed study of the two-dimensional case, addressing also generalized FEM to reduce the pollution error, will be presented in a forthcoming paper.

Figure 26: Values  $C_1$  computed from (39)

## 4 Conclusions

1. Numerical solutions to Helmholtz problems do not preserve, in general, the nondispersive property of the physical medium and the exact mathematical solution. While exact solutions in the form of propagating waves travel with a velocity that depends on material constants of the medium only, both the velocity and even the form of the numerical solution depend significantly on the parameter  $k$  of the Helmholtz equation. A propagating numerical solution displays a phase velocity that is a function of  $k$ .

In addition to the dependence on the physical parameter  $k$ , the numerical wave also depends on the numerical parameters  $h$  (stepsize) and  $p$  (polynomial order of approximation). From the viewpoint of practical application, dispersion analysis and numerical analysis pursue a common goal: *a priori* assessment of the quality and reliability of the numerical solution in terms of the physical parameter  $k$  and the numerical parameters  $h$  and  $p$ .

2. In dispersion analysis, quality and reliability of the numerical solution are related to its phase accuracy. Thus, the phase difference between the exact and the numerical solution is assessed analytically and by numerical experiment on model problems. The result is expressed as a function of  $k, h$  and  $p$ . It is found that, for any fixed  $p$ , the phase difference grows with wavenumber  $k$  also on normalized (i.e. the product  $hk$  is

constrained) mesh. In other words, the phase difference between the exact (nondispersive) and the numerical (dispersive) solution is itself dispersive. This is shown analytically for Galerkin finite element solutions to one-dimensional Helmholtz problems.

For the  $h - p$ -method, the notation of normalization is generalized to a constraint of the scale  $hk/p$ , thus controlling the ratio of the wavenumber  $k$  and the number of DOF,  $ph^{-1}$ . While the Galerkin FEM solution is dispersive for any  $p$ , the phase accuracy improves, for fixed scale  $\theta = hk/p < 1$  and fixed  $k$ , significantly with growing  $p$ . More precisely, the error can be estimated from above by a term proportional to  $(\theta * e/4)^{2p}/p^{1/2}$ . This result analytically confirms former conclusions from numerical experiments. It is shown by numerical evaluation in this paper that the upper estimate of phase accuracy given in Theorem 1 is sharp, i.e. not generally improvable.

3. Numerical analysis of the Helmholtz equation leads to error estimates in integral norms. Previously known estimates had shown that the relative numerical error, measured in  $H^1$ -norm, for small stepsize  $h$  depends on the scale of the mesh only, but not on the wavenumber  $k$ . Expanding the analysis to the practically relevant stepsizes  $h$  as obtained by normalization w.r. to the wavenumber  $k$ , one obtains a different result. Namely, it is shown (both theoretically and by numerical experiment) that the relative numerical error on normalized mesh contains a term of the order of the phase lead. Hence, the numerical error in integral norm is dispersive on normalized mesh. The theoretical upper estimates are confirmed by numerical evaluation. Dispersion analysis and numerical analysis, while using different methods of investigation, thus lead to equivalent conclusions concerning the quality and reliability of the numerical solution.

4. For the one-dimensional model problems considered by the present as well as by other authors, both the phase accuracy and the numerical accuracy in integral norm improve with increasing  $p$ . More precisely, if the wavenumber  $k$  is normalized by the number of DOF (by means of fixing the  $h - p$ -scale), then for any fixed  $k$  the error decreases exponentially with increase of  $p$ . Relatively most significant improvement is obtained when passing from the standard  $h$ -version (with  $p = 1$ ) of the Galerkin FEM to approximation order  $p = 2$  or  $p = 3$ .

The improvement of phase accuracy achieved by stabilization methods is thus less significant for higher  $p$  as compared to the standard version with  $p = 1$ .

Similar phenomena to the ones discussed here have been observed in one-dimensional problems of fluid-structure interaction and two-dimensional problems. The observations made so far lead to the conjecture that the theoretical results on the error behaviour, as obtained for one-dimensional problems, carry over to the two-dimensional case. While a two-dimensional example is evaluated here, results on fluid-structure interaction and further analysis of two-D problems will be presented in forthcoming papers.

**Acknowledgement:** This work was supported by ONR Grant N00014-93-I-0131. The first author was supported partly by Grant No 517 402 524 3 of the German Academic



Exchange Service (DAAD). The computations for the two-dimensional example were carried out by Ms. Ellen Paik, Computer Science Dept., University of Maryland at College Park

## References

- [1] Abboud, N. N.; Pinsky P.M.: Finite element dispersion analysis for the three-dimensional second-order scalar wave equation. *Int. J. Numer. Met. Eng.* 35 (1992), 1183-1218
- [2] Aziz, A.K., Kellogg R.B., Stephens, A.B.: A two point boundary value problem with a rapidly oscillating solution, *Numer. Math.* 53, 107-121 (1988)
- [3] Babuška, I. and Aziz, A.K.: The mathematical foundations of the finite element method, in: A.K. Aziz (ed.), *The mathematical foundations of the finite element method with applications to partial differential equations*, Academic Press, New York 1972, pp. 5-359
- [4] Babuška, I.; Ihlenburg, F.; Makridakis, Ch.: Analysis and finite element methods for a fluid solid interaction problem in one dimension, Technical Note, Institute for Physical Science and Technology, University of Maryland at College Park (in preparation)
- [5] Babuška, I.; Katz, I.N.; Szabó, B.S.: Finite element analysis in one dimension, Lecture Notes, to appear in Springer Verlag
- [6] Babuška, I.; Sauter, S.: Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers, Technical Note BN-1172 (1994), Institute for Physical Science and Technology, University of Maryland at College Park
- [7] Bayliss, A.; Goldstein, C.I.; Turkel, E.: On accuracy conditions for the numerical computation of waves, *J. Comp. Phys.* 59 (1985), 396-404
- [8] Ciarlet, P.G.: *The finite element method for elliptic equations*, North Holland 1978
- [9] Demkowicz, L.: Asymptotic convergence in finite and boundary element methods: part I: theoretical results, preprint 1993
- [10] Douglas Jr., J.; Santos, J.E.; Sheen, D.; Schreiyer, L.: Frequency domain treatment of one-dimensional scalar waves, *Mathematical Models and Methods in Applied Sciences*, Vol. 3, No. 2 (1993) 171-194
- [11] Harari, I. ; Hughes, T.J.R.: Finite element method for the Helmholtz equation in an exterior domain: model problems, *Comp. Meth. Appl. Mech. Eng.* 87 (1991), 59-96
- [12] Harari, I. ; Hughes, T.J.R.: A cost comparison of boundary element and finite element methods for problems of time-harmonic acoustics, *Comp. Meth. Appl. Mech. Eng.* 97 (1992), 77-102
- [13] Harari, I. ; Hughes, T.J.R.: Galerkin/least squares finite element methods for the reduced wave equation with non-reflecting boundary conditions in unbounded domains, *Comp. Meth. Appl. Mech. Eng.* 98 (1992) 411-454

- [14] Ihlenburg F.; Babuška, I.: Finite element solution to the Helmholtz equation with high wavenumber - part I: The h-version of the FEM, Technical Note BN-1159 (1993), Institute for Physical Science and Technology, University of Maryland at College Park
- [15] Ihlenburg F.; Babuška, I.: Finite element solution to the Helmholtz equation with high wavenumber - part II: The h-p-version of the FEM, Technical Note BN- (1994), Institute for Physical Science and Technology, University of Maryland at College Park
- [16] Schatz, A.H.: An observation concerning Ritz-Galerkin methods with indefinite bilinear forms, *Math. Comp.* 28 (1974), 959-962
- [17] Schatz, A.H.: An analysis of the finite element method for second order elliptic boundary value problems, in: A.H. Schatz, V.T. Thomée and W. L. Wendland, *Mathematical theory of finite and boundary element methods*, Birkhäuser Verlag 1990
- [18] Szabó B.; Babuška, I.: *Finite Element Analysis*, J. Wiley, New York etc., 1991
- [19] Thompson L.L.; Pinsky, P.M.: Complex wavenumber Fourier analysis of the  $p$ -version finite element method, *Computational Mechanics*, 13 (1994), 255 -275
- [20] Thompson L.L.; Pinsky, P.M.: A Galerkin Least Squares Finite Element Method for the Two-Dimensional Helmholtz Equation, *Int. J. Num. Meth. Engng.* (to appear)
- [21] Thompson L.L.: Design and analysis of space-time and Galerkin/ least squares finite element methods for fluid structure interaction in exterior domains Ph.D. Dissertation, Stanford University, April 1994

**The Laboratory for Numerical Analysis is an integral part of the Institute for Physical Science and Technology of the University of Maryland, under the general administration of the Director, Institute for Physical Science and Technology. It has the following goals:**

**To conduct research in the mathematical theory and computational implementation of numerical analysis and related topics, with emphasis on the numerical treatment of linear and nonlinear differential equations and problems in linear and nonlinear algebra.**

**To help bridge gaps between computational directions in engineering, physics, etc., and those in the mathematical community.**

**To provide a limited consulting service in all areas of numerical mathematics to the University as a whole, and also to government agencies and industries in the State of Maryland and the Washington Metropolitan area.**

**To assist with the education of numerical analysts, especially at the postdoctoral level, in conjunction with the Interdisciplinary Applied Mathematics Program and the programs of the Mathematics and Computer Science Departments. This includes active collaboration with government agencies such as the National Institute of Standards and Technology.**

**To be an international center of study and research for foreign students in numerical mathematics who are supported by foreign governments or exchange agencies (Fulbright, etc.).**

**Further information may be obtained from Professor I. Babuška, Chairman, Laboratory for Numerical Analysis, Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742-2431.**